CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 126, No. 5 November 2009

JASA EXPRESS LETTERS

LETTERS TO THE EDITOR

GENERAL LINEAR ACOUSTICS [20]

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

Ultra-wide sensor arcs for low frequency sonar detection with a baffled cylindrical array

Derek C. Bertilone, Chaoying Bao, Ben C. Travaglione, and Damien S. Killeen
Defence Science and Technology Organisation, Building A51, HMAS Stirling, P.O. Box 2188, Rockingham DC, Western Australia 6958, Australia
derek.bertilone@dsto.defence.gov.au, chaoying.bao@dsto.defence.gov.au, ben.travaglione@dsto.defence.gov.au, damien.killeen@dsto.defence.gov.au

Abstract: Passive detection with a baffled cylindrical array can potentially be improved at low frequencies by exploiting signal diffraction around the baffle. A model based on infinite rigid cylinder scattering suggests that large gains in signal-to-noise ratio are potentially available to adaptive beamformers if the sensor arc is widened to include sensors in the acoustic shadow. However, elastic scatter effects become increasingly important as frequency decreases, so the gains obtained in practice are unknown. The gains in detection performance are examined in this letter by analyzing data recorded at sea from a platform-mounted sonar array.

PACS numbers: 43.60.Fg [JC]
Date Received: June 1, 2009 **Date Accepted:** July 24, 2009

1. Introduction

Passive cylindrical sonar arrays are operated over the widest possible frequency range to exploit all available acoustic energy. But it is a challenge to obtain acceptable performance at low frequencies due to high levels of ambient and platform-generated noises and poor bearing resolution.[1] Adaptive beamformers (ABFs) are important for this purpose. In addition to providing improved bearing resolution and side-lobe suppression, ABFs can potentially provide a higher array gain than conventional beamformers if the noise has a high degree of spatial correlation.[2] This is the case for ambient noise at low frequencies.[1] Typically, the beamformer processes an arc of sensors mounted on a cylindrical metal baffle. An arc smaller than 180° is typically used because sensors in the acoustic shadow of the baffle have low signal-to-noise ratio (SNR) and make little contribution to the output over most of the frequency range. At low frequencies, however, significant signal energy is diffracted around the baffle, and detection may be improved by using a larger arc.

Because ABF is sensitive to steering vector errors,[2] the ability to exploit these additional gains depends on accurate modeling of the signal in the acoustic shadow region. The usual approach is to treat the baffle as a rigid scatterer.[3,4] Meyer,[5] Teutsch and Kellermann,[6] Teutsch,[7] and Bertilone *et al.*[8] analyzed beamforming, detection, localization, and array gain for arrays mounted on rigid baffles. However, the rigid model is inadequate at low frequencies where the elastic properties of the materials are important.[4] Unfortunately it is difficult to develop a more accurate model, as it requires detailed analysis of the baffle, its mounting to the platform, and scattering from other parts of the platform. Despite the limitations of the infinite rigid cylinder model, it is of interest to examine the gains achievable in practice when it is incorporated into an ABF with ultra-wide arc. This letter presents results obtained using data recorded at sea from a platform-mounted array, as the arc is extended into the acoustic shadow by increasing the number of processed sensors while keeping sensor spacing fixed.

2. Signal model and ABF

The array is mounted at the front of a platform and has Q staves uniformly spaced on a circle of radius r, surrounding a cylindrical metal baffle of radius $a \leq r$. Each stave is a line of omnidirectional phones parallel to the cylinder axis. To simplify the discussion, we analyze data where the phones in each stave have been summed with zero time-delay. Thus the array can be viewed

as a baffled circular array of directional sensors, in which each sensor is a line array steered to broadside. Signals are assumed to be plane-waves at zero-elevation, i.e., perpendicular to the cylinder axis. The beamformer processes $M \leq Q$ sensors that lie on an arc that swings around with the steering direction. In practice, the acoustic signals from distant sources often exhibit small deviations from zero-elevation, leading to phase errors, but these errors are minor at the low frequencies considered in this letter.

Modeling the baffle as an infinite rigid cylinder, and ignoring scatter from other parts of the platform, we introduce cylindrical coordinates (r, ϕ, z) with z-axis at the center of the baffle. If a plane wave signal of unit amplitude and wavenumber $k=2\pi f/c$ arrives normal to the z-axis from azimuth $\phi=0°$, then the complex acoustic field outside the baffle is[3]

$$\psi(r,\phi) = \exp(-jkr\cos\phi) + \sum_{n=0}^{\infty} \varepsilon_n (-j)^n b_n H_n^{(1)}(kr)\cos(n\phi), \tag{1}$$

$$b_n = -\frac{J_{n-1}(ka) - J_{n+1}(ka)}{H_{n-1}^{(1)}(ka) - H_{n+1}^{(1)}(ka)}. \tag{2}$$

Here $H_n^{(1)}$ and J_n are Hankel and Bessel functions of the first kind, respectively, $\varepsilon_0=1$ and $\varepsilon_n=2$, $n \geq 1$. If the array is steered to $\phi=0°$ using sensors at $(r,\phi_1)\cdots(r,\phi_M)$, then the steering vector is $\boldsymbol{\psi}=[\psi(r,\phi_1)\cdots\psi(r,\phi_M)]^T$, where T denotes transpose. For frequency-domain beamforming,[2] complex sensor outputs at frequency f are obtained by fast Fourier transformation (FFT) and placed in an $M\times 1$ vector $\mathbf{X}$. Doing this for I snapshots of data, we form

$$\hat{\mathbf{R}} = \frac{1}{I}\sum_{i=1}^{I} \mathbf{X}^{(i)}\mathbf{X}^{(i)H}, \tag{3}$$

as an estimate the cross-spectral matrix (CSM), where H denotes conjugate transpose. Output power is

$$P = \mathbf{v}^H \hat{\mathbf{R}} \mathbf{v}, \tag{4}$$

where $\mathbf{v}$ is the weight vector. We use a widely studied ABF, the minimum power distortionless response beamformer with sample matrix inverse[2] (MPDR SMI), for which

$$\mathbf{v} = \frac{\hat{\mathbf{R}}^{-1}\boldsymbol{\psi}}{\boldsymbol{\psi}^H \hat{\mathbf{R}}^{-1}\boldsymbol{\psi}}. \tag{5}$$

The effect of baffle scatter on the amplitude and phase of $\boldsymbol{\psi}$ has been discussed elsewhere.[8]

The top row of Fig. 1 shows power vs bearing near a contact, computed from trials data using MPDR SMI with $\boldsymbol{\psi}$ constructed using the infinite rigid cylinder model. The data were processed using FFTs with 32 Hz bins and Hann windowing. The CSM was estimated using 96 snapshots with 50% overlap. No diagonal loading[2] was used. Outputs are shown for arcs of sizes 112.5°, 180°, and 225° at normalized frequencies $ka=3.1, 4.5, 6.1$. We observe a large improvement in bearing resolution and noise suppression as the arc is extended from 112.5° to 180°, but also a noticeable improvement as the arc is extended into the acoustic shadow from 180° to 225°. Bearing resolution for ABF is improved as the arc extends beyond 180° because it is not determined solely by the physical aperture and frequency.

3. SNR and detection index

We analyze the change in output SNR as the size of the arc is increased from 112.5° to Φ, $\Delta\mathrm{SNR}(\Phi)=\mathrm{SNR}(\Phi)-\mathrm{SNR}(112.5°)$. Here $\mathrm{SNR}=10\log_{10}[(P_{S+N}-P_N)/P_N]$, where P_{S+N} is the mean output power when signal and noise are present, and P_N is the mean output power when noise-only is present. The dashed curves in the middle row of Fig. 1 show predictions of ΔSNR using an array gain model that treats the baffle as an infinite rigid cylinder, and the noise field as

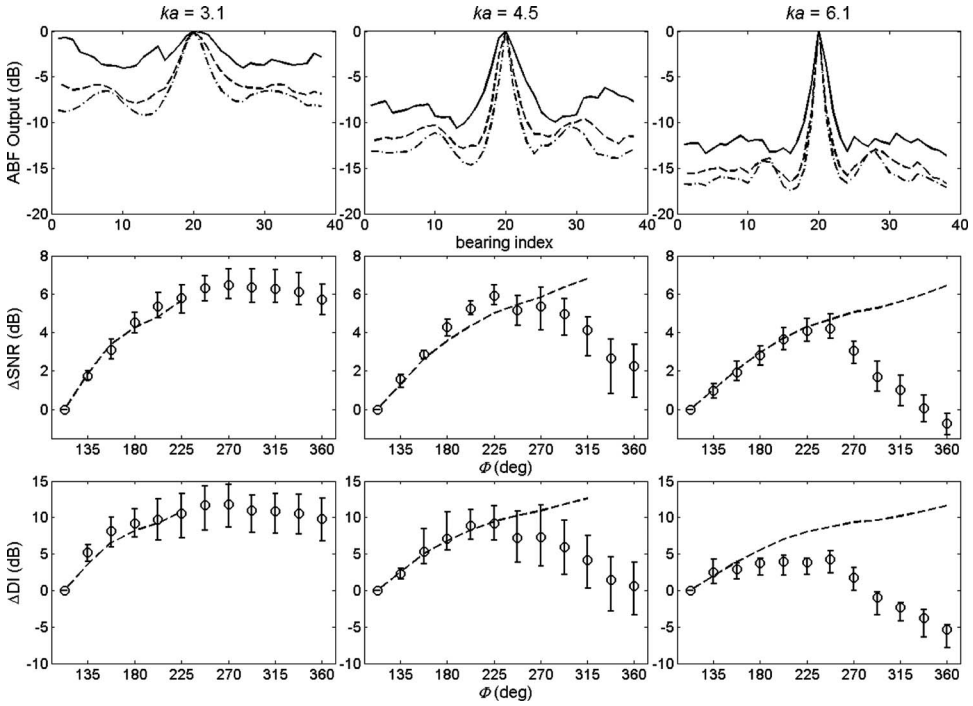


Fig. 1. Top: Power vs bearing near a contact, from applying ABF to trial data using arcs of sizes 112.5° (solid), 180° (dashed), and 225° (dash-dot). Middle: Change in output SNR as a function of arc size. Circles with error bars show experimental results using ABF, and dashed curves show model predictions. Bottom: Change in DI as a function of arc size. Circles with error bars show experimental results using ABF, and dashed curves show model predictions.

a superposition of independent plane-waves with surface dipole power distribution.[8] Results are shown for optimum processing to maximize SNR.[2] The surface dipole model is often used to represent ambient noise in deep water originating from wave action at the surface;[1] the noise arrives from above the array (i.e., at elevations $0° \leq \theta \leq 90°$) with power per unit steradian proportional to $\sin\theta$. The modeling suggests that large increases in SNR are potentially available. At $ka = 4.5$, for example, SNR is increased by almost 4 dB as the arc increases from 112.5° to 180°, and by a further 3 dB when it increases from 180° to 315°. Note that the modeling requires inversion of a noise CSM and could not be obtained for all arcs at the lowest frequencies due to ill-conditioning of the matrix.

To examine the gains obtained in practice, we computed 12 estimates of ΔSNR for each arc, by processing consecutive segments of the same data recording used in the top row of Fig. 1, using the same processing parameters. SNR was estimated by replacing P_{S+N} by the power at the contact bearing, and P_N by the average power over a window of bearings that excluded the signal. The circles in the middle row of Fig. 1 show the mean value of the estimates, and error bars indicate upper and lower quartiles so that 50% of the estimates lie within the indicated bounds. Note that if $\boldsymbol{\psi}$ was, in fact, the true steering vector, and if the sample CSM was the true CSM, then MPDR SMI would have a mean output that achieves the maximum SNR when steered to a solitary signal in noise.[2] In fact, for arcs up to 247.5° or 270° we do indeed find that the experimental SNRs are in broad agreement with the model prediction. However, SNR drops away as the arc is extended further. We find a SNR increase of 3–4 dB when the arc increases from 112.5° to 180°, and up to 2 dB of additional gain when it is increased from 180° to a value between 225° and 270°. The drop-off in SNR occurs because ABF is sensitive to steering vector errors,[2] and these errors grow as sensors deep inside the acoustic shadow are included in the processing. The errors are most likely due to a combination of elastic

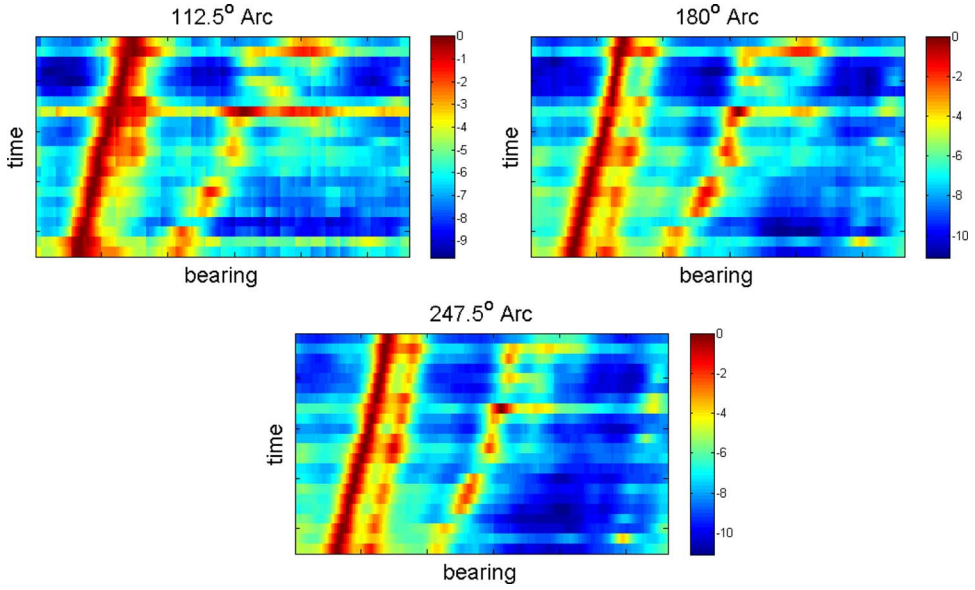


Fig. 2. (Color online) Broadband bearing-time records for data processed using ABF with arcs of various sizes. Four contacts can be distinguished.

scatter effects, and scattering from other parts of the platform. ABF can be made more robust by diagonally loading[2] the CSM in Eq. (5), but we found that this did not allow larger arcs to be utilized.

The bottom row of Fig. 1 shows the change in detection index[9] (DI) as the arc is increased from 112.5°, $\Delta\mathrm{DI}(\Phi)=\mathrm{DI}(\Phi)-\mathrm{DI}(112.5°)$. Here $\mathrm{DI}=20\log_{10}[(P_{S+N}-P_N)/\sigma_N]$, where σ_N is the standard deviation of the output power when noise-only is present. DI is more directly related to detection performance than SNR because it explicitly accounts for fluctuation in the background noise. Capon and Goodman[10] showed that $\sigma_N=P_N/\sqrt{(I-M+1)}$ if certain statistical properties of the noise field are applicable, and this leads to $\mathrm{DI}=2\mathrm{SNR}+10\log_{10}(I-M+1)$. Dashed curves show predictions obtained by applying this formula with the surface dipole noise model. Circles with error bars show experimental results obtained by applying ABF to the data and replacing σ_N by the sample standard deviation of power in a window of bearings that excludes the signal. At the lower frequencies the model is in good agreement with the data for arcs up to the optimum size, but at the higher frequency the model and data diverge almost immediately. The cause of the rapid divergence is unclear, but in this case there is little to be gained by extending the arc into the shadow zone. At the two lower frequencies we find large increases in DI of 7–9 dB as the arc is increased from 112.5° to 180°, and an additional increase of more than 2 dB as the arc is increased to its optimum size. To put these numbers into context, we utilize the receiver operating characteristic $P_d=\mathrm{erfc}[\mathrm{erfc}^{-1}(P_f)-10^{\mathrm{DI}/20}]$ for signal detection from Gaussian distribution theory[9] when the signal is sufficiently weak that the noise-only variance approximates the signal plus noise variance. Here P_d and P_f are the detection and false alarm probabilities, respectively, and erfc is the complementary error function. Assuming that this approximates the true detection statistics when the number of snapshots is sufficiently large, then a DI of 9.8 dB allows a signal to be detected 50% of the time at $P_f=0.1\%$. An additional 2 dB raises P_d to 79%, while a reduction of 2 dB lowers it to 26%.

4. Bearing-time record

The gains in a multiple contact scenario are illustrated in Fig. 2, which shows broadband power displayed as a bearing-time record, computed from trials data using ABF with the same param-

eters as used in Fig. 1. Background noise was equalized[9] before power was summed over the band $ka=3-6$. Horizontal scans were normalized to a maximum of 0 dB. The image at top left shows the output for a 112.5° arc. Two contacts are observed: a strong contact at left and a weaker contact near the center. The contact at left is broad, suggesting that it might be comprised of multiple contacts that cannot be spatially resolved. The image at top right shows the output for a 180° arc. The contact at left can now be resolved into a strong contact and nearby weak contact. A fourth contact at far right can only just be discerned. The bottom image shows the output for a 247.5° arc. There are noticeable improvements in the quality of the tracks, which are primarily due to increases in signal to interference plus noise ratio for the weak signals.

5. Conclusion

We have demonstrated that low frequency sonar detection with a cylindrical array can be improved by exploiting signal diffraction around the baffle. ABF was applied to data recorded at sea from a platform-mounted array, using a steering vector constructed from the infinite rigid cylinder model. Best results were obtained using sensor arcs in the range 225°–270°, depending on the frequency, but further increases caused a drop-off in performance due to growing steering vector errors. Additional gains could be accessed if a better signal model was available or if an experimentally measured steering vector was used.

References and links

[1]W. S. Burdic, *Underwater Acoustic System Analysis*, 2nd ed. (Prentice-Hall, Englewood Cliffs, NJ, 1991).
[2]H. L. Van Trees, *Optimum Array Processing* (Wiley-Interscience, New York, 2002), Pt. IV.
[3]E. A. Skelton and J. H. James, *Theoretical Acoustics of Underwater Structures* (Imperial College Press, London, 1997).
[4]M. C. Junger and D. Feit, *Sound, Structures, and Their Interaction*, 2nd ed. (Acoustical Society of America, Melville, NY, 1986).
[5]J. Meyer, "Beamforming for a circular microphone array mounted on spherically shaped objects," J. Acoust. Soc. Am. **109**, 185–193 (2001).
[6]H. Teutsch and W. Kellermann, "Acoustic source detection and localization based on wavefield decomposition using circular microphone arrays," J. Acoust. Soc. Am. **120**, 2724–2736 (2006).
[7]H. Teutsch, *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition* (Springer, Berlin, 2007).
[8]D. C. Bertilone, D. S. Killeen, and C. Bao, "Array gain for a cylindrical array with baffle scatter effects," J. Acoust. Soc. Am. **122**, 2679–2685 (2007).
[9]R. O. Nielsen, *Sonar Signal Processing* (Artech House, Boston, MA, 1991).
[10]J. Capon and N. R. Goodman, "Probability distributions for estimators of the frequency-wavenumber spectrum," Proc. IEEE **58**, 1785–1786 (1970).

Explosion localization via infrasound

Curt A. L. Szuberla, John V. Olson, and Kenneth M. Arnoult
Wilson Infrasound Observatories, Geophysical Institute, University of Alaska Fairbanks, Fairbanks, Alaska 99775-7320
cas@gi.alaska.edu, jvo@gi.alaska.edu, kma@gi.alaska.edu

Abstract: Two acoustic source localization techniques were applied to infrasonic data and their relative performance was assessed. The standard approach for low-frequency localization uses an ensemble of small arrays to separately estimate far-field source bearings, resulting in a solution from the various back azimuths. This method was compared to one developed by the authors that treats the smaller subarrays as a single, meta-array. In numerical simulation and a field experiment, the latter technique was found to provide improved localization precision everywhere in the vicinity of a 3-km-aperture meta-array, often by an order of magnitude.

PACS numbers: 43.60.Jn, 43.28.Dm [VO]
Date Received: July 18, 2009 **Date Accepted:** August 11, 2009

1. Introduction

A problem for military ground forces is the localization of explosions in 10×10 km^2 areas of operation, often of the improvised explosive device (IED) type, to a precision of order 10 m. Acoustic methods of localization have long been used for this purpose because of their relatively inexpensive nature. Detonations of more than a kilogram of conventional explosive represent a low frequency acoustic source, with energy primarily in the infrasound frequency band ($f < 20$ Hz). In this frequency band, the traditional localization method is called data fusion, or BAZ in this paper. The BAZ technique uses an ensemble of arrays to separately estimate direction-of-arrival (DOA) information, resulting in a localization solution estimated from the various back azimuths. The DOA estimates are said to be far field, since the aperture of each subarray is assumed to be small compared to the source distance. This technique has application across a broad range of interest, from nuclear treaty monitoring,[1] to vehicle tracking,[2] to the conventional detonations[3,4] described in this study. DOA-based methods of localization are prone to uncertainties arising from atmospheric,[5] environmental,[6] and intrinsic[7] factors. Taken together, these account for a precision of order 100 m in practical applications consistent with the aim of this study.[4]

In the literature there is a certain paucity of infrasound localization applications at ranges less than 100 km. Part of this stems from the difficulty in applying high-resolution techniques to the data, which are often contaminated by wind noise and create difficulty in forming simple signal models. For infrasonic localization, spectral-estimation-based methods are not useful, due to significant departures from $1/r$ pressure fluctuations.[8,9] Experience with applying techniques that directly estimate wavefront curvature to infrasound data with known ground truth for near-field sources has shown that these methods are similarly not useful for infrasonic applications. Neither BAZ nor srcLoc suffer from these limitations.

Acoustic localization across 10×10 km^2 areas may also be accomplished by employing near-field assumptions. Near-field methods variously make use of DOA and/or time-difference of arrival information (TDOA).[10–12] Efforts to apply these two techniques to infrasonic data led the authors to develop a near field, strictly TDOA-based method of acoustic localization, or srcLoc in this paper. This technique treats each of the subarrays of the BAZ method as part of a single, meta-array. While all TDOA-based methods can be shown to represent the optimal intersection of hyperbolic curves in a phase space,[8] application of the srcLoc method to a wide variety of synthetic and actual infrasound signals has shown it to outperform other near-field techniques. Mathematically, the srcLoc method calculates an optimal, in some

sense (typically least squares), intersection of sensor world lines with a source sound cone in position-velocity-time space.[9] An analytic least squares solution of the cone intersections serves as a seed for a numerical optimization routine. For infrasound localization, the advantages of the srcLoc method lie primarily in the absence of restrictive atmospheric assumptions. The atmosphere is assumed, albeit unrealistically, to be isotropic and windless, leading to a right, circular sound cone. Too, there is no implicit model assumption governing the functional form of the signal source; since only TDOA information needs to be estimated, the isotropic atmosphere assumption is sufficient.

This paper describes a numerical simulation and field experiment to test the relative localization efficacy of BAZ and srcLoc on infrasound data. The goal of the experiments was twofold. First, to determine if the newly developed srcLoc technique would yield enhanced localization precision over the traditional BAZ method in this particular application, and second, to determine if the simple model assumptions behind srcLoc would hold up under scrutiny in the field.

2. Numerical simulation

The sponsor of this study placed a broad constraint on the application: an area of 10×10 km^2 should be covered by a dozen sensors for the purpose of low-frequency acoustic localization of explosions. In part, this constraint stems from past practice and budget limitations. Prior experience with portable infrasound array deployments lead to an array design of three subarrays, each comprising four sensors. A subarray consisted of sensors positioned at the vertices of a square, 100 m from a central point (what would become the digitizer location). The centroids of each subarray were then positioned at the vertices of an equilateral triangle, 3 km on a side (the meta-array). Such a design would give adequate spatiotemporal resolution for both the BAZ and srcLoc techniques to provide localization across much of the area.

The effects of 500 blasts at the center of each 100×100 m^2 pixel in a 144 km^2 area were simulated. In practice, the TDOA information required for BAZ and srcLoc is estimated from generalized cross correlation of sampled waveforms. This estimation process is relatively slow, so this study made use of synthetic TDOA information representative of 10 dB signal to noise ratio (SNR) blasts when sampled at 1 kHz. Specifically, perfectly sampled TDOA information for each blast was contaminated by the appropriate (empirically determined) amount of Gaussian noise. For BAZ, the TDOA information leads to three DOA estimates (one for each subarray), from which a localization solution is calculated. For srcLoc, a single localization solution is calculated (via an analytic seed fed to a Nelder–Mead optimization[13] routine) from the entire TDOA ensemble.

The distribution of absolute range errors (δ_R) was estimated for each pixel and technique. The value of δ_R corresponding to a cumulative sum of 0.95 on the distribution was mapped to a color, as shown in Fig. 1. No pixel in the simulation area has a larger error for srcLoc than for BAZ. By using srcLoc in lieu of the traditional BAZ, an improvement of roughly an order of magnitude is expected throughout much of the simulation area. Everywhere in the interior of the meta-array, srcLoc is predicted to exhibit the desired precision of order 10 m. The BAZ technique gives rise to large range errors when the DOA estimate from any one cluster is directed toward another. This problem is greatly exaggerated when a blast occurs outside the meta-array. Because of this, the results depicted in Fig. 1 were clipped at 6 km errors for the BAZ technique.

3. Experiment

From 27–29 August 2007, an experiment was conducted on the range complex of Ft. Greely, Alaska, in order to verify the predictions detailed in Sec. 2. The actual emplacement of instruments and demolitions is depicted in Fig. 2. Each position was surveyed via 10-min averaged GPS data, to a precision of ±3 m. While the vagaries of terrain and vegetation qualitatively distorted the planned array geometry, the predictions of Sec. 2 still apply. Infrasound sensors used in the experiment were Chaparral Physics Mod. 25 and each cluster of four was sampled at 1 kHz. For wind-noise reduction, clusters were placed in patches of boreal forest (small, dense

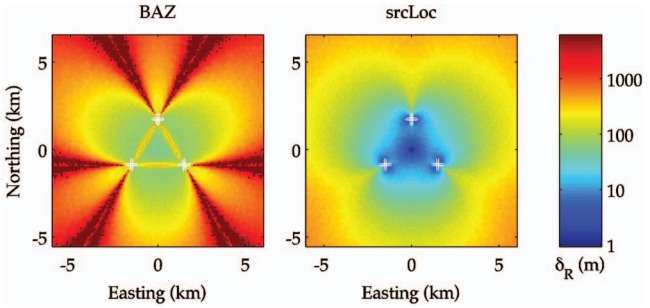


Fig. 1. Confidence limits (95%) for the absolute range errors in simulated source localization for each technique. The traditional DOA method of data fusion (BAZ) is depicted in the left panel and that of the strictly TDOA technique (srcLoc), in the right. To construct the panels, an ensemble of 500 TDOA vectors for each 100×100 m^2 pixel was synthesized. The TDOA information was representative of 10 dB SNR acoustic arrivals across the array when sampled at 1 kHz. The distributions of absolute range errors were estimated for each pixel and technique. The 95% confidence limits in the range errors are mapped to color in each panel [the left (BAZ) panel was clipped at 6 km errors]. White + symbols represent sensor locations.

trees and brush) and each sensor was connected to four, 20 m porous hoses from various (unknown) manufacturers. No meteorological data were acquired, consistent with the sponsor's application; however, the weather was sunny and calm on both days. Observations of drifting smoke and dust from the detonations indicated that the winds were variable (direction) and less than about 2 m/s (walking speed).

Demolitions were positioned at the sites depicted in Fig. 2. These represent a compromise between testing various critical locations in the plots of Fig. 1 and satisfying U.S. Army training requirements. Additionally, Ft. Greely Range Control criteria had to be met as the fire danger is typically high during August in interior Alaska. Explosives used in the experiment were blocks (0.57 kg) of M112 C-4 plastic explosive. The explosives were variously staked at

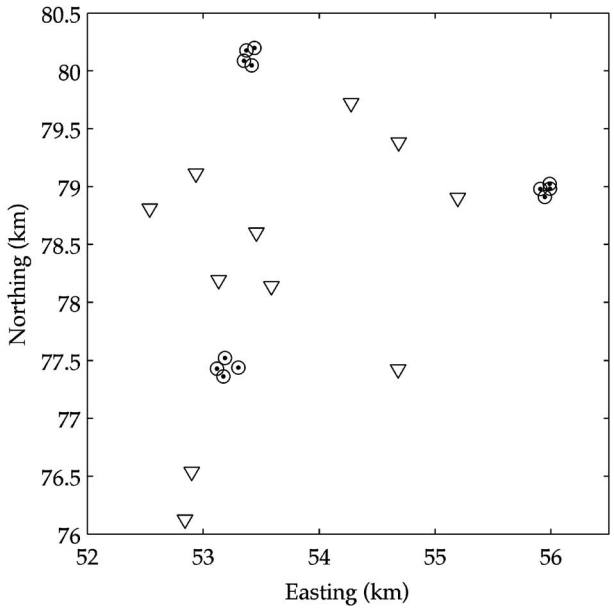


Fig. 2. Experimental array and blast sites. Sensor locations (dotted circles) and blast sites (triangles) are depicted. Distinct charges of M112 C-4, in various sizes, were detonated at each site (see text for details). Coordinates (km) are in grid 06VWR of the MRGS coordinate system.

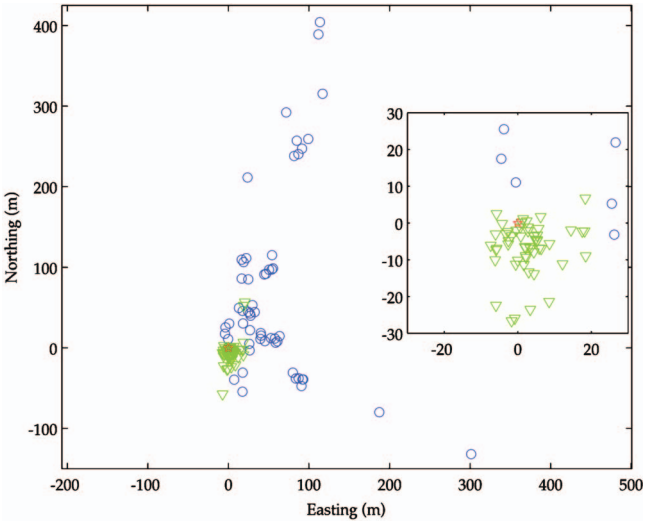


Fig. 3. Normalized results of experimental source localizations for each technique. Each blast site is translated to the origin (red star) and the same translation is applied to the corresponding localization solutions (blue circles for BAZ and green triangles for srcLoc). The inset depicts a 30×30 m^2 box about the origin, in which 95% of the srcLoc solutions lie, but only 11% of the BAZ solutions.

1 m height or placed directly on the ground. Initiation was via hand-pulled time fuse and blasting caps. Ground surface ranged from dry, hard-pack dirt road to wet, muskeg bog. Blast sites were also variously in the open or sheltered by boreal forest. At each site between three and seven separate charges, ranging in size from one-quarter to four blocks, were detonated. This amounted to a total of 55 separate blasts.

Although spectral localization methods were not employed, each of the blasts exhibited a broad peak in energy at roughly 18 Hz, across all of the sensors. Thus the data were bandpass filtered at $f \in [0.8, 36]$ Hz; however, the results were not sensitive to the exact corner frequencies. This filtering resulted in further noise reduction and is a standard part of infrasonic data processing.[1] TDOA information for the subarrays (BAZ) and the meta-array (srcLoc) was then estimated via cross correlation. The TDOA information alone was used to localize each blast with srcLoc. In the case of BAZ, DOA estimates were made prior to localization. Each blast site was translated to the origin and the same translation was applied to each respective localization solution. These results are depicted in Fig. 3.

Of the 55 blasts, srcLoc produced a range error $\delta_R \leqslant 30$ m for 95% of them, compared to 11% using BAZ (as seen in the inset of Fig. 3). The srcLoc method gave better than a factor of 2 increase in localization precision over BAZ at 96% of the sites, and in no instance less than a factor of 1.4. An order of magnitude increase was obtained with srcLoc at 50% of the blast sites.

4. Conclusions

In assessing the goal of the experiment, srcLoc was found everywhere to provide enhanced localization precision over that of BAZ, both in simulation and in the field. Additionally, the simple atmospheric assumptions that underlie srcLoc were sufficient to achieve this enhancement during the testing period. A difference in the bias of each localization method was noted, but is not explained by the underlying model assumptions. Further study of this phenomenon is required. An inadvertent test of sensitivity to sensor failure was conducted as indigenous hares gnawed on the cabling at the eastern subarray, disabling a sensor for a number of blasts. The effect of sensor loss was also studied by purposefully ignoring the data from particular sensors. Single sensor failures such as this will have a large impact on a four-element subarray, mani-

festing itself as an increase in the uncertainty in one of the DOA estimates that feeds BAZ.[7] This single-sensor failure mode has virtually no effect on srcLoc. Under srcLoc, there is no requirement for small clusters of sensors; with sufficient digitizers, sensors can be randomly scattered throughout the area of interest. That said, clusters are a convenient deployment strategy in case atmospheric conditions become severe enough to distort TDOA information at long (metaarray), but not short (subarrays), ranges. BAZ then can serve as a backup for srcLoc in a practical setting.

Beyond the scope of military applications, the technique can be applied to geophysical situations. Near-field volcano monitoring is an example of current interest, where enhanced, low-frequency acoustic localization can tag a particular vent as being active or lead to the location of a new fumarole.

Acknowledgments

This work represents one aspect of the research supported by NSF Grant No. IIS-0433392 and Geophysical Institute internal funding. This experiment would not have been possible were it not for the support of J. Helmericks of the Geophysical Institute, SFC Wolter and the soldiers of the 73rd ENGR CO, 1-25 SBCT, Ft. Wainwright, AK. The authors thank Dr. D. Withoff for helpful discussions and the reviewers for their cogent and insightful comments.

References and Links

[1]J. V. Olson and C. A. L. Szuberla, "Processing infrasonic array data," in *Handbook of Signal Processing in Acoustics*, edited by D. Havelock, S. Kuwano, and M. Vorländer (Springer, New York, 2008), Vol. **2**, pp. 1487–1496.

[2]R. J. Kozick and B. M. Sadler, "Source localization with distributed sensor arrays and partial spatial coherence," IEEE Trans. Signal Process. **52**, 601–616 (2004).

[3]B. G. Ferguson, L. G. Criswick, and K. W. Lo, "Locating far-field impulsive sound sources in air by triangulation," J. Acoust. Soc. Am. **111**, 104–116 (2002).

[4]V. Pinsky, Y. Gitterman, A. Hofstetter, and A. Shapira, "Robust location of surface explosions by a network of acoustic arrays," Geophys. Res. Lett. **33**, L02317 (2006).

[5]D. K. Wilson, "Performance bounds for acoustic direction-of-arrival arrays operating in atmospheric turbulence," J. Acoust. Soc. Am. **103**, 1306–1319 (1998).

[6]B. G. Ferguson and K. W. Lo, "Passive ranging errors due to multipath distortion of deterministic transient signals with application to the localization of small arms fire," J. Acoust. Soc. Am. **111**, 117–128 (2002).

[7]C. A. L. Szuberla and J. V. Olson, "Uncertainties associated with parameter estimation in atmospheric infrasound arrays," J. Acoust. Soc. Am. **115**, 253–258 (2004).

[8]J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays: Signal Processing Techniques and Applications*, edited by M. Brandstein and D. Ward (Springer, New York, 2001), pp. 157–180.

[9]C. A. L. Szuberla, K. M. Arnoult, and J. V. Olson, "Performance of an infrasound source localization algorithm," J. Acoust. Soc. Am. **120**(5), 3179 (2006).

[10]B. G. Ferguson, "Variability in the passive ranging of acoustic sources in air using a wave-front curvature technique," J. Acoust. Soc. Am. **108**, 1535–1544 (2000).

[11]T. Duong Tran-Luu and P. Cremona-Simmons, "Acoustic sources localization by simultaneous DOA and TOA," in *Proceedings of the 2006 Meeting of the Military Sensing Symposia (MSS) Specialty Group on Battlespace Acoustic & Seismic Sensing, Magnetic & Electric Field Sensors (BAMS)* (US Army CERDEC Night Vision and Electronic Sensors Directorate, Ft. Belvoir, VA, 2007), pp. 441–452.

[12]C. A. L. Szuberla, K. M. Arnoult, and J. V. Olson, "Discrimination of near-field infrasound sources based on time-difference of arrival information," J. Acoust. Soc. Am. **120**, EL23–EL28 (2006).

[13]W. A. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. (Cambridge University Press, New York, 1992).

The use of non-collinear mixing for nonlinear ultrasonic detection of plasticity and fatigue

Anthony J. Croxford,[a] Paul D. Wilcox, and Bruce W. Drinkwater
Department of Mechanical Engineering, University of Bristol, Bristol BS8 1TR, United Kingdom
a.j.croxford@bristol.ac.uk, p.wilcox@bristol.ac.uk, b.drinkwater@bristol.ac.uk

Peter B. Nagy
Department of Aerospace Engineering, University of Cincinnati, Cincinnati, Ohio 45221
peter.nagy@uc.edu

Abstract: This letter reports on the application of the non-collinear mixing technique to the ultrasonic measurement of material nonlinearity to assess plasticity and fatigue damage. Non-collinear mixing is potentially more attractive for assessing material state than other nonlinear ultrasonic techniques because system nonlinearities can be both independently measured and largely eliminated. Here, measurements made on a sample after plastic deformation and on a sample subjected to low-cycle fatigue show that the non-collinear technique is indeed capable of measuring changes in both, and is therefore a viable inspection technique for these types of material degradation.

PACS numbers: 43.35.Zc, 43.35.Yb, 43.25.Zx [JM]
Date Received: July 29, 2009 **Date Accepted:** August 26, 2009

1. Introduction

Nonlinear ultrasonic measurements enable the detection of the onset of plastic deformation and fatigue damage at an earlier stage than conventional linear nondestructive testing (NDT) techniques, which have insufficient sensitivity to the changes in the microstructure brought on by dislocation movements. Finite-deformation elastic theory introduces three independent constants, referred to as third order elastic constants (TOECs), which describe the nonlinear stress-strain behavior in an isotropic material.[1,2] Different sets of independent TOECs have been proposed by various authors, including A, B, and C used by Landau and Lifshitz,[1] which are a linear combination of the l, m, and n Murnaghan constants.[2]

Of practical interest is the dependence of TOECs on the level of plastic strain or fatigue damage induced dislocation accumulation in a material. Various ultrasonic methods of measuring material nonlinearity have been developed. The first makes use of the so-called acousto-elastic effect.[3,4] In this case the nonlinear behavior manifests itself through variations in ultrasonic propagation velocity with applied strain. Through the application of different wave types and the measurement of velocity in unstrained and strained states all three TOECs can be measured. One problem with this technique is the difficulty of measuring the small changes in propagation time and distance accurately enough to allow the velocity, and from that the TOECs, to be determined. A second problem is the necessity of loading a specimen to measure the changes in velocity.

The second and perhaps most widely reported method for interrogating material nonlinearity is the harmonic generation technique.[5–8] If ultrasonic energy at one frequency is injected into a material, harmonics of the input frequency are generated due to nonlinearity as the ultrasound propagates. By measuring the magnitude of the harmonics the degree of material nonlinearity can be quantified. There is a considerable body of experimental evidence that

[a]Author to whom correspondence should be addressed.

shows a strong correlation between the normalized harmonic amplitude and the amount of fatigue damage[6] or plastic deformation[7] in a material. The major measurement difficulty with the harmonic generation method as a NDT technique lies in isolating the causes of nonlinearity. Specifically, amplifiers, transducers, and coupling methods are all contributors to the measured harmonic, often on a scale greater than the material nonlinearity itself. Thus it is practically very difficult to determine if the measured nonlinearity is due to the material or the equipment.

A third technique, which is the main subject of this paper, for TOEC measurement was first proposed by Jones and Kobett,[9] and experimentally observed by Rollins.[10] This approach is based on the fact that material nonlinearities cause interaction between two intersecting ultrasonic waves.[11] Under certain circumstances, this can lead to the generation of a third wave with a frequency and wavevector equal to the sum of the incident wave frequencies and wavevectors, respectively. Theoretically, there are several incident wave combinations that can achieve this; however, practical material constraints to the theory lead to the interaction of two shear waves generating a longitudinal wave as the most useful case.

The non-collinear mixing technique has two important advantages over the conventional nonlinear ultrasonic harmonic generation technique. First, it is much less sensitive to system nonlinearities due to spatial selectivity (the nonlinear interaction is limited to the region where the incident beams intersect), modal selectivity (the nonlinear mixing signal is a different mode to the incident waves), frequency selectivity (the mixing signal frequency can be separated from harmonics of the incident waves if the driving frequencies are chosen to be unequal), and directional selectivity (the mixing signal propagates in a different direction form the mixed ones and their higher harmonics). Second, unlike the harmonic generation techniques, the level of the underlying system nonlinearity can be measured directly by summing the responses to each of the incident waves excited separately, that is, without the interaction present.

It is important to note that the evidence of correlation between material degradation (e.g., fatigue or plasticity) and nonlinear ultrasonic phenomena that has been reported is based mainly on evidence from the harmonic generation technique. In this configuration, only longitudinal waves can be used, and the harmonic amplitude is a function of all three TOECs (A, B, and C) or alternatively two of Murnaghan's three TOECs (l and m). However, the non-collinear technique based on the interaction of two shear waves to produce a longitudinal wave was shown by Jones and Kobett[9] and Taylor and Rollins[11] to lead to a longitudinal wave amplitude that depends only on TOECs A and B (or the m and n Murnaghan TOECs).

What has not been studied to date is whether the particular combination of the two TOECs probed by the non-collinear technique is sensitive to fatigue and plasticity, and therefore whether the non-collinear technique can be used for NDT of fatigue damage. The purpose of this letter is to demonstrate that the non-collinear mixing technique can indeed detect changes due to plasticity and fatigue damage, and therefore has the potential to be used as a NDT technique.

2. Experimental arrangement

Experimental measurements were performed on an Al2014-T4 aluminum alloy specimen. Figure 1 shows the basic experimental arrangement. Two intersecting shear waves are generated using oblique incidence shear transducers made of longitudinal transducers of 5 MHz nominal center frequency mounted on 60° Perspex wedges. Within the volume of intersection a third longitudinal wave is generated due to nonlinear interaction. Once generated, this wave propagates through the material in a conventional manner and is detected by the receiver. The receiver was a normal-incidence longitudinal transducer of 10 MHz nominal center frequency. The excitation signals were generated using a digital oscilloscope/signal generator and the detected interaction wave (after amplification) was recorded using the same instrument. The excitation signals were amplified using a power amplifier, resulting in signals with amplitude of approximately 60 $V_{p\text{-}p}$.

The excitation signals to both input transducers were 20-cycle, Hanning-windowed tone bursts with center frequencies of 5.5 MHz. Using the same driving frequency for both incident waves removes one of the advantages of the non-collinear technique (frequency sepa-

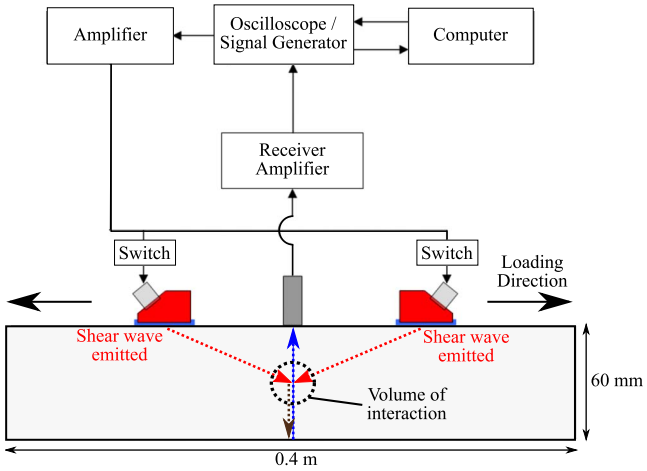


Fig. 1. (Color online) Experimental arrangement.

ration) although the following results show that ample suppression of system nonlinearities is still achieved. The recorded data were digitally filtered using an 11 MHz center frequency, 2 MHz bandwidth bandpass filter. The use of incident waves of the same frequency significantly simplifies the experimental apparatus as only one common driving signal needs to be generated. It also simplifies the experimental geometry since with both incident waves excited at equal and opposite angles, the resulting interaction wave is generated perpendicular to the specimen surface. The latter point means that the receiver can be placed on either the top or bottom surface of the specimen. For the purpose of this investigation the single sided arrangement was considered more suitable as it reflects the limited access likely to be encountered in practical applications. Note that the vertical position of the interaction zone can be readily moved by altering the separation of the input transducers.

Throughout all stages of experimentation each test comprised three measurements: one with each input transducer excited individually and one with both excited simultaneously. The signals recorded when the input transducers were excited individually were summed and the amplitude of this signal at the expected arrival time of the interaction wave used to estimate the level of remnant system nonlinearity. Figure 2(a) shows an example of time-domain response obtained from an as-manufactured sample when both input transducers are excited simultaneously. The pulse in the window labeled first reflection is the first received interaction wave after it has been reflected off the bottom surface of the sample. The subsequent pulses in the windows labeled second and third reflections correspond to reverberations of the interaction wave between the sample surfaces. In the following, the peak amplitude in the window corresponding to the first reflection is taken as the measure of material nonlinearity. Figure 2(b) shows the equivalent time-domain signal obtained by summing the responses from each of the two input transducers excited separately. The amplitude of the signal in the window labeled first reflection in Fig. 2(b) is due to the combined effect of all nonlinearities in the measurement system (e.g., reflections due to sidelobes of harmonics in the transmitted shear waves). It can be seen that the system nonlinearity is an order of magnitude smaller than the material nonlinearity, yielding a signal-to-noise ratio of 30, even in the as-manufactured sample, which is expected to contain the lowest nonlinearity anyway.

It is worth noting the presence of the signal at 2.5×10^{-5} S in Fig. 2(a) resulting from second harmonic generation on a longitudinal wave propagating through the specimen and off of the back wall. If a conventional harmonic generation technique were employed this signal would be impossible to differentiate from any potential equipment nonlinearity, whereas using the non-collinear technique the interaction wave is spatially separated.

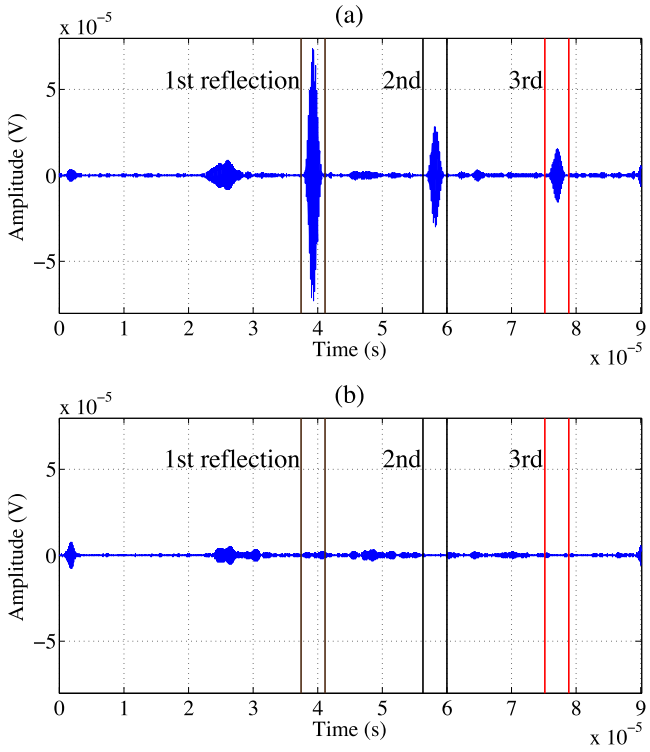


Fig. 2. (Color online) Time-domain signals obtained from as-manufactured sample corresponding to (a) the total response when the shear transducers are excited simultaneously and (b) the sum of the responses when the shear transducers are excited separately. The colored lines show the calculated arrival times of the first, second, and third reflections.

The responses of the three transducers used were calibrated to absolute values using a heterodyne laser interferometer. The purpose of this calibration was to enable a theoretical value of interaction wave amplitude for the as-manufactured material to be estimated for comparison with that measured experimentally. Values of -25.22×10^{10}, -32.5×10^{10}, and -35.12×10^{10} $\mathrm{N/m^2}$ for the TOECs for single crystal aluminum were taken from literature.[12] The approximate amplitude of the interaction wave was then estimated using the expression (Table I, case I) provided by Taylor and Rollins,[11] yielding a value of 3.2×10^{-12} m. The measured amplitude of the interaction wave for the intact sample was 2×10^{-12} m. This is sufficiently close to the estimated value to give confidence to the basic soundness of the non-collinear measurement technique. The difference is believed to be primarily due to the theoretical values being calculated for a single crystal and therefore not taking into account the polycrystalline nature of the real sample. This calibration procedure illustrates the means by which the non-collinear technique could be used for making absolute measurements.

Having confirmed that the measured nonlinear interaction wave was of similar amplitude to that predicted theoretically, experiments were carried out to investigate changes in the magnitude of the interaction wave as the material was subjected to both quasi-static plastic strain and low-cycle fatigue damage. The first sample was used to investigate the effect of plastic deformation. Strain gauges were bonded to opposing faces of the sample directly above the region of interaction and the sample was loaded in a tensile test machine. After removing the load, the residual plastic strain was measured using the strain gauges and the specimen was removed from the test machine for the non-collinear measurements to be performed. Each set of non-collinear measurements corresponded to eight measurement points along the length of the

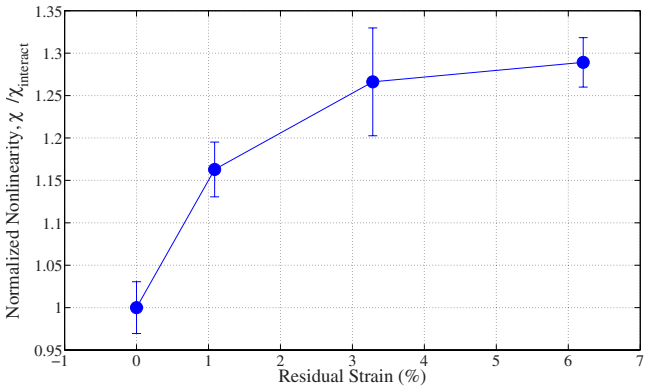


Fig. 3. (Color online) Change in non-collinear measurement parameter χ with increasing plastic strain.

sample. The process was repeated using the same specimen subjected to progressively higher loads to obtain non-collinear mixing data at successively higher levels of plastic strain.

In order to measure the material nonlinearity independent of the excitation level the amplitude of the interaction wave A_3 was normalized to the product of the two input amplitudes A_1 and A_2 measured in volts.

$$\chi = \frac{A_3}{A_1 A_2}. \tag{1}$$

Figure 3 shows the measured values of χ as a function of residual strain normalized to the intact value in order to make the change clearer. The key point to observe in this graph is the increase in χ by around 30% with residual strain, indicating that the non-collinear approach is sensitive to plasticity. Each point on the graph represents the mean of the eight individual measurements made for that particular plastic strain level. Between each measurement the transducer fixture was completely removed and the specimen cleaned. The error bars represent one standard deviation of the measurements.

A second specimen was tested under low-cycle fatigue conditions. These correspond to cyclically straining the material to beyond its yield point but significantly below its failure stress. Typical fatigue life under these conditions is less than 100 cycles. In the example presented here the sample was stressed between 0% and 110% of yield (420 MPa) in blocks of 10 cycles. Initially, this stress level led to a residual strain of 2%, significantly below the failure strain, but still high enough to result in a low-cycle fatigue failure. The results of this test are shown in Fig. 4.

It can be seen that χ initially increases rapidly with the number of cycles. Beyond 20 cycles the rate of increase drops significantly due to work-hardening in the material, and this is in line with published data in literature.[8] In this experiment the error bars get larger with increasing number of cycles indicating a higher degree of variability in the measurements. This can be attributed to taking measurements along the whole of the test section rather than at a single location. End effects near the points of attachment to the tensile test machine may well result in more localized fatigue damage, hence increasing the variability of measurements.

3. Conclusions

These results demonstrate that the non-collinear technique is sensitive to both plasticity and fatigue damage in a similar way to collinear harmonic generation. This is despite the non-collinear technique being sensitive to a combination of different TOECs to the harmonic generation technique. Because of its intrinsically better rejection of spurious system nonlinearities,

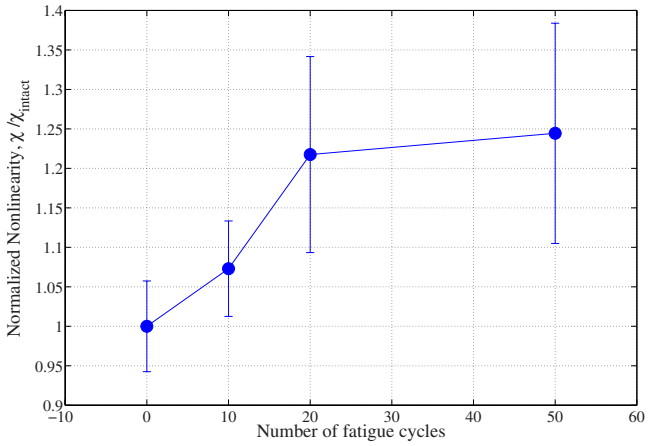


Fig. 4. (Color online) Change in non-collinear measurement parameter χ with increasing loading cycles.

the non-collinear mixing technique is highly suitable for measurement of weak material nonlinearity, which can be exploited in the future for robust NDT of service-related damage in critical components.

References and links

[1]L. D. Landau and E. M. Lifshitz, *Theory of Elasticity* (Pergamon, New York, 1959).
[2]F. D. Murnaghan, "Finite deformations of an elastic solid," Am. J. Math. **59**, 235–260 (1937).
[3]R. A. Toupin and B. Bernstein, "Sound waves in deformed perfectly elastic materials—Acoustoelastic effect," J. Acoust. Soc. Am. **33**, 216–225 (1961).
[4]P. B. Nagy, "Fatigue damage assessment by nonlinear ultrasonic materials characterization," Ultrasonics **36**, 375–381 (1998).
[5]J. H. Cantrell and K. Salama, "Acoustoelastic characterisation of materials," Int. Mater. Rev. **36**, 125–145 (1991).
[6]J. H. Cantrell and W. T. Yost, "Nonlinear ultrasonic characterization of fatigue microstructures," Int. J. Fatigue **23**, 487–490 (2001).
[7]J. Kim, L. J. Jacobs, J. Qu, and J. W. Littles, "Experimental characterization of fatigue damage in a nickel-base superalloy using nonlinear ultrasonic waves," J. Acoust. Soc. Am. **120**, 1266–1273 (2006).
[8]C. Pruell, J. Kim, J. Qu, and L. J. Jacobs, "Evaluation of plasticity driven material damage using Lamb waves," Appl. Phys. Lett. **91**, 231911 (2007).
[9]G. L. Jones and D. R. Kobett, "Interaction of elastic waves in an isotropic solid," J. Acoust. Soc. Am. **35**, 5–10 (1963).
[10]F. R. Rollins, "Interaction of ultrasonic waves in solid media," Appl. Phys. Lett. **2**, 147–148 (1963).
[11]L. H. Taylor and F. R. Rollins, Jr., "Ultrasonic study of three-phonon interactions. I. Theory," Phys. Rev. **136**, A591–A596 (1964).
[12]J. H. Cantrell, "Crystalline structure and symmetry dependence of acoustic nonlinearity parameters," J. Appl. Phys. **76**, 3372–3380 (1994).

Attention modulates auditory adaptation produced by amplitude modulation

Takayuki Kawashima
Graduate School of Health Sciences, Teikyo-Heisei University, 2-51-4 Higashi-Ikebukuro, Toshima-ku, Tokyo 170-8445, Japan
takayuki.kawashima@thu.ac.jp

Abstract: The effect of attention on adaptation produced by amplitude modulation (AM) was examined. In different experimental conditions, listeners' AM detection thresholds for a 2 kHz test tone were measured after exposing them to an adapting sound that was presented simultaneously with speech distractors. Magnitude of an aftereffect, calculated as the elevation of the thresholds caused by adaptation, was smaller when the listeners shift attention away from the adaptor to the distractor voice than when they attended to the adaptor. The results suggest that the AM of unattended sounds may not be fully analyzed compared to that of attended sounds.

PACS numbers: 43.66.Mk, 43.66.Rq [QJF]
Date Received: June 29, 2009 **Date Accepted:** August 24, 2009

1. Introduction

Almost all sounds in natural environments, including our voices, have an amplitude envelope that varies over time. This time-dependent variation [amplitude modulation (AM)] plays important roles in auditory perception, as is seen in speech perception (e.g., Shannon *et al.*, 1995) and in perceptual grouping of frequency components (Darwin and Carlyon, 1995). To understand the processing of amplitude envelopes, several researchers have examined characteristics of perceptual adaptation produced by AM. For example, Kay and Matthews (1972) reported that the AM detection threshold increased following listeners' exposure to a long, modulated tone (i.e., adaptor). Wojtczak and Viemeister (2003) used a matching procedure to show that suprathreshold perception of modulated tones was also changed by adaptation. Moreover, to some extent, such aftereffects of AM exposure appear to be selectively linked to both the rate of modulation and the carrier frequency of the adaptor sound (Kashino, 1998; Richards *et al.*, 1997; Wojtczak and Viemeister, 2003). To some, such findings suggest that aftereffects of AM adaptation arise from specific auditory processing of AM (e.g., Wojtczak and Viemeister, 2003).

An important problem related to adaptation by AM concerns the involvement of attention. Does auditory processing show constant adaptation irrespective of the direction of the listener's attention? In vision, several studies have reported that adaptation is indeed affected by attention (e.g., Yeh *et al.*, 1996). For example, Chaudhuri (1990) examined the effect of attention on the motion aftereffect elicited by a moving array of random dots. The duration of the aftereffect was shorter when the observers shifted attention from the dots to alphanumeric symbols superimposed on the dots, compared to when they attended to the moving dots.

In the present report, the author examined possible effects of attention on adaptation by AM. Following research in visual perception, the author predicted that AM adaptation should be less pronounced when the listener's attention is directed to sounds other than the adaptor sounds during an adapting phase than when attention is focused directly on the adaptors.

2. Method

2.1 Participant

Nine observers, two males and seven females, with no reported history of hearing problem, participated in the experiment. They were university students and paid for the participation.

2.2 Stimuli and equipment

Sound stimuli were generated on a personal computer (Apple, Macintosh G4) equipped with a sound card and digital to analog converter. They were presented to individual participants over headphones (Sennheiser, HDA200). All stimuli were digitized in 16 bits and were played at 44.1-kHz sample frequency. Measurements were conducted in a sound-proof room.

A sinusoidally amplitude-modulated (SAM) tone with a carrier frequency of 2 kHz was served in both the test stimuli and the adaptor. Two types of AM test stimuli were created based, respectively, on two different modulation frequencies: 16 and 51 Hz. Durations of the test (SAM) tones were always 0.5 s. The modulation frequency of the adaptor was 16 Hz; its depth of modulation was −3 dB. To prepare for the manipulation of the listener's attention, the level of the adaptor was occasionally changed during presentation. The fluctuation of sound level lasted for 1 s (during this interval, the sound was damped linearly for 0.4 s, then kept at minimum sound level of −8 dB for 0.2 s, and was ramped linearly for 0.4 s to the original level). The time interval between fluctuations was randomly chosen from three candidates: 1.4, 2.4, and 3.4 s. The adaptor and the test stimuli were always presented through the left channel.

Three types of distractor sounds were used: female voices, male voices, and 5313-Hz tone. Distractors were presented simultaneously with an adaptor sound. The voice distractors were always presented at the opposite ear from the adaptor. The female voice distractor consisted of utterances of numbers (1–10 in Japanese, and the average duration was 0.32 s); ordering of numbers was randomized with replacement. The utterances were separated by silent intervals whose duration varied randomly among three values: 0.50, 0.60, and 0.65 s. The male voice distractor was composed of sentences in a speech corpus (Speech Database Committee of Acoustical Society of Japan, 1990); this continuous distractor was included to render a focus on numbers (in the female voice) more difficult. Both voices were filtered with a band-stop filter to reduce their energy around 2 kHz. Unlike voice distractors, the tone distractor was presented through the same channel as the adaptor. The rationale for adding a tone distractor was to lower adaptor salience. It has been reported that listeners have difficulty in attending to one of the two sounds simultaneously presented at one ear, when other sound was presented at the opposite ear (Kidd *et al.*, 2003). All stimuli were presented at a sound pressure level of 61 dB, although each voice distractor was presented at 60 dB. All stimuli were gated on and off with 0.02 s cosinusoidal amplitude ramps.

2.3 Procedure

Thresholds for AM detection were measured using a two-interval forced choice procedure. Equally often a SAM tone occurred in one of two 0.5 s intervals; a silent 0.3 s interval separated two observation intervals. Participants reported which interval contained a SAM tone. Feedback was not presented (e.g., Richards *et al.*, 1997). In conditions where an adaptor was presented, a long adaptor (80 s duration) was presented at the start of a block of trials. In addition, shorter adaptors (40 s durations) were presented between trials when 25 s have passed since the last presentation of the adaptors.

In addition to a no-adaptor control condition, three adapting conditions were examined. In the *adaptor* condition, adapting stimuli were presented accompanied only by the tone distractor. In two different adaptor-plus-voices conditions, the adaptor and all three types of the distractors were simultaneously presented; these two conditions were acoustically identical; however, they differed in attentional set and the listeners were imposed on attentional tasks during adaptor presentation (in addition to AM detection). One of the two conditions, the *attend-to-adaptor* condition, required participants to report the level fluctuations of the adaptor by pressing a key during the adaptor presentation. The other condition was the *attend-to-distractor* condition. In this condition, participants were required to judge (by pressing one of two keys) whether an utterance of the female voice expressed a number that was even or odd.

All participants received the control condition and the adaptor condition. However, they received only one of the two adaptor-plus-voices conditions; five were assigned to the attend-to-adaptor condition and four were assigned to the attend-to-distractor condition. Since

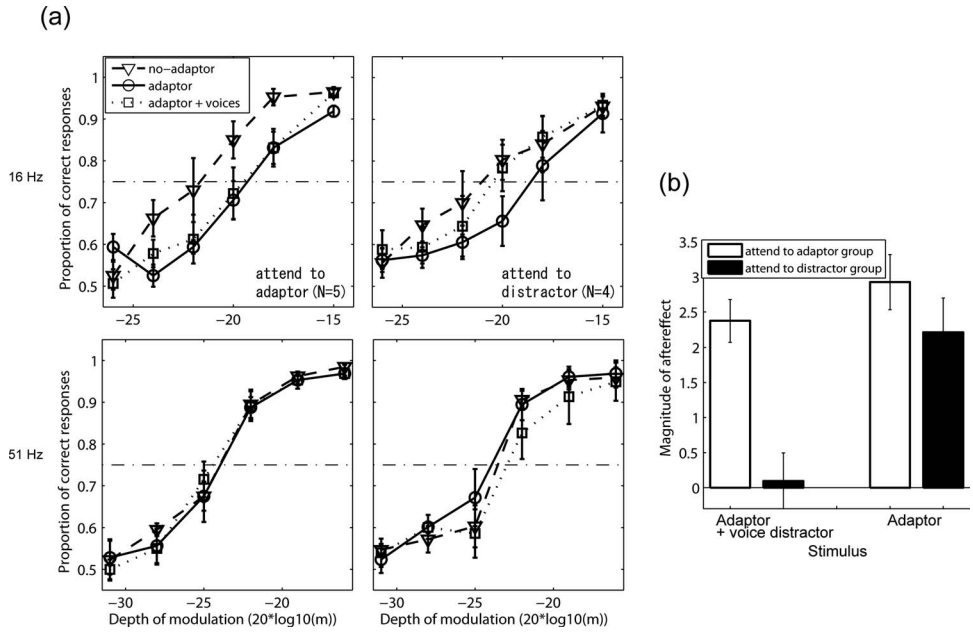


Fig. 1. (a) Psychometric functions of the attend-to-adaptor group (left side panels) and the attend-to-distractor group (right side panels) for two different test modulation frequencies (16 and 51 Hz). (b) Magnitude of aftereffects. In both of the panels, in the adaptor condition, the two participant groups were exposed to the same adapting setup (adaptor plus a tone distractor) and were not imposed on attentional tasks during the adaptor presentation. All error bars reflect standard errors. See text for details.

both these groups received the same adapting stimuli but different attentional instructions, the design of the experiment was a $2 \times 3 \times 2$ mixed factorial design in which two levels of test modulation rates and three levels of the adapting settings (no-adaptor, adaptor with a tone distractor, and adaptor with three distractors) were crossed with two levels of the instructions.

A trial block contained 96 trials (2 modulation rates $\times$ 6 depths of modulation $\times$ 8 iterations); a block typically lasted 10 min in conditions that presented the adaptor. Each participant completed eight blocks for each condition (in the attend-to-distractor condition, one participant with time constraints received only seven blocks). A single session consisted of two blocks of the control plus (following) four blocks of either the adaptor condition or the adaptor-plus-voices condition. Different adapting conditions (i.e., the adaptor and the adaptor-plus-voices conditions) were alternately tested in separate days, their start conditions being counterbalanced across participants, following a first session that contained practice blocks.

3. Results and discussion

Psychometric functions for each of the two attention groups were plotted using the average data points [Fig. 1(a)]. In both groups, the functions in the 16 Hz modulation condition (upper row) shifted to the right when the adaptor was presented (circles) compared to the no-adaptor control (triangles). The shift indexes the aftereffect due to AM adaptation.

When the distractor voices were presented, reduction in the performance by the adaptation was softened at several tested modulation depths in the attend-to-distractor group (squares at upper right panel). To compare the effect of the adaptor across conditions, the magnitude of the aftereffect was calculated for each participant as a threshold difference between an adapted condition and that in the control condition (the threshold used corresponds to the modulation depth associated with 75% correct responding, specified by the cumulative normal functions fitted using least squares). The average magnitude in the 16 Hz test condition, plotted in Fig. 1(b), shows effects of attention condition and adaptor. When distractor voices were not

presented, the magnitude of an AM aftereffect was almost constant in both the attend-to-distractor group and the attend-to-adaptor group (two bars at the right side). However, when distractor voices were presented, the magnitude was less pronounced in the attend-to-distractor group than in the attend-to-adaptor group (two bars at the left side). A 3×2 analysis of variance (ANOVA) using the threshold values (for the 16 Hz condition, the results of two factor analysis, not three, were reported for description brevity) revealed a significant interaction between the adaptor conditions and attention instructions, $F(1,7)=6.47$, $p<0.01$. Multiple comparisons, using Tukey's honestly significant differences tests, indicated that the thresholds in the adaptor condition increased significantly relative to those in the no-adaptor control condition ($p<0.01$). This increment was not modulated by the presences of distractor voices in the attend-to-adaptor group ($p>0.1$); however, the threshold decreased in the attend-to-distractor group ($p<0.01$). These results indicate that adaptation becomes less pronounced when listeners shift attention away from the adaptor.

When the modulation frequency of the adaptor (16 Hz) differed from that of the test stimuli (51 Hz), the proportion of correct responses did not largely differ across conditions [Fig. 1(a), bottom rows]. As before, a 3×2 ANOVA using the thresholds was conducted. No statistically significant effects of the adaptation variable, the attention variable, and their interaction were found ($p>0.1$ in all cases). The results are in line with the above-mentioned conjecture that the change in the magnitude of the aftereffect in 16 Hz condition is due to the shift of attention, not to the general influence of the distractor task.

Because the number of presentations of adapting stimulus varied over blocks depending on the responding pace of each participant, it is possible that results of Fig. 1 were influenced by this variable. However, we found no systematic differences in the number of the adaptor presentations between the attend-to-adaptor group and the attend-to-distractor group (8.4 and 8.7 on average, in the adaptor-plus-voices condition, respectively). This indicates that the number of the adaptor presentation cannot explain the difference in the magnitude of the aftereffect.

Accuracy of performance in two distractor tasks was relatively high throughout the experiment. When attending to the distractor voice, the proportion of correct number identifications was between 0.87 and 0.91 across the listeners. In the attend-to-adaptor condition, the hit rate for detecting a level fluctuation in the adaptor ranged between 0.76 and 0.96, with false alarms being infrequent (8.3 times per block, on average). These results suggest that the listener's attention was manipulated as expected by the instructions.

The observed effects of attention on AM adaptation suggest that AM of an unattended sound is not fully analyzed relative to the level of AM analysis conferred on an attended sound. In natural environments, where the auditory system must solve the difficult problem of sound segregation, it may be beneficial for the system to relax analysis of AM of unattended sounds at some stage of processing. By this, system resources can be concentrated on segregating the target sound. Indeed, the effect of attention observed in the present study might relate to the strategy that the auditory system has adopted for efficient processing.

Bruckert *et al.* (2006) reported that training reduced the AM adaptation. From the finding, they discussed a possibility that the aftereffect might reflect listener's misuse of the adaptor as a reference for detection (see also Wakefield and Viemeister, 1984). In the present study, although an effort was taken to reduce the similar misuse, the distractors might prevent the listener's use of an incorrect reference and this might explain some of the reduction in the aftereffect.

Although attention affects AM adaptation, the exact nature of attention involved in the present experiment remains unclear. There are several possible interpretations of its nature. The first of which is that the number judgment task required large cognitive load, which in turn would lead to disruption of AM processing. Another possibility is that attending to the distractor voice may hinder the processing of the stimuli presented to the contralateral side, leading to less pronounced adaptation. It would be helpful to examine the nature of attention related to the observed effect for further understanding of the auditory processing of AM.

Acknowledgments

This work was supported by KAKENHI (Grant No. 19730456) and 21st Century COE programs "Center for Evolutionary Cognitive Sciences," MEXT, Japan.

References and links

Bruckert, L., Herrmann, M., and Lorenzi, C. (**2006**). "No adaptation in the amplitude modulation domain in trained listeners," J. Acoust. Soc. Am. **119**, 3542–3545.

Chaudhuri, A. (**1990**). "Modulation of the motion aftereffect by selective attention," Nature (London) **344**, 60–62.

Darwin, C., and Carlyon, R. P. (**1995**). "Auditory grouping," in *Hearing*, edited by B. C. J. Moore (Academic, London).

Kashino, M. (**1998**). "A modulation filter bank as a basis for auditory spectro-temporal analysis," Fifth International Conference on Neural Information Processing, Vol. **3**, pp. 1333–1336.

Kay, R. H., and Matthews, D. R. (**1972**). "On the existence in human auditory pathways of channels selectively tuned to modulation present in frequency-modulated tones," J. Physiol. (London) **225**, 657–667.

Kidd, G., Jr., Mason, C. R., Arbogast, T. L., Brungart, D. S., and Simpson, B. D. (**2003**). "Informational masking caused by contralateral stimulation," J. Acoust. Soc. Am. **113**, 1594–1603.

Richards, V. M., Buss, E., and Tian, L. (**1997**). "Effects of modulator phase for comodulation masking release and modulation detection interference," J. Acoust. Soc. Am. **102**, 468–476.

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

Speech Database Committee of Acoustical Society of Japan (**1990**). "ASJ-JIPDEC speech corpus."

Wakefield, G. H., and Viemeister, N. F. (**1984**). "Selective adaptation to linear frequency-modulated sweeps: Evidence for direction-specific FM channels?," J. Acoust. Soc. Am. **75**, 1588–1592.

Wojtczak, M., and Viemeister, N. F. (**2003**). "Suprathreshold effects of adaptation produced by amplitude modulation," J. Acoust. Soc. Am. **114**, 991–997.

Yeh, S. L., Chen, I. P., De Valois, K., and De Valois, R. L. (**1996**). "Figural aftereffects and spatial attention," J. Exp. Psychol. Hum. Percept. Perform. **22**, 446–460.

Perceptual fusion of polyphonic pitch in cochlear implant users

Patrick J. Donnelly
Department of Computer Science, Johns Hopkins University, Baltimore, Maryland 21218
donnell@cs.jhu.edu

Benjamin Z. Guo and Charles J. Limb
Department of Otolaryngology—Head and Neck Surgery, Johns Hopkins Hospital, Baltimore, Maryland 21287
bguo@fas.harvard.edu, climb@jhmi.edu

Abstract: In music, multiple pitches often occur simultaneously, an essential feature of harmony. In the present study, the authors assessed the ability of cochlear implant (CI) users to perceive polyphonic pitch. Acoustically presented stimuli consisted of one, two, or three superposed tones with different fundamental frequencies (f_0). The normal hearing control group obtained significantly higher mean scores than the CI group. CI users performed near chance levels in recognizing two- and three-pitch stimuli, and demonstrated perceptual fusion of multiple pitches as single-pitch units. These results suggest that limitations in polyphonic pitch perception may significantly impair music perception in CI users.

PACS numbers: 43.66.Ts, 43.66.Hg, 43.75.Cd [QJF]
Date Received: July 29, 2009 **Date Accepted:** August 27, 2009

1. Introduction

The ability of cochlear implant (CI) users to perceive music remains severely limited, primarily due to limited pitch resolution (Gfeller *et al.*, 2002; Pijl, 1997). While studies have shown that CI users are able to adequately perceive temporal cues that convey rhythmic information, the perception of pitch and timbre remains quite poor with current CI hardware and processing strategies (Leal *et al.*, 2003; McDermott, 2004). Previous studies have shown that CI users are severely impaired compared to normal hearing (NH) subjects in tests of pitch perception using acoustically presented stimuli, with CI users rarely exhibiting pitch discrimination thresholds of less than several semitones (Gfeller *et al.*, 2002; Looi *et al.*, 2004). Most published studies on pitch perception have focused on pitch discrimination, in which subjects are required to detect whether two sounds differ in pitch, and pitch ranking, in which subjects are asked to listen to two sounds presented in sequence and judge which one has the higher pitch. While these approaches are certainly valid, various elements (i.e., melody, harmony, rhythm, and timbre) usually occur simultaneously in music. In the context of the impaired pitch resolution described in CI users, it is germane that nearly all forms of music utilize at least some degree of polyphony (where multiple pitches occur simultaneously), an essential feature of harmony. Comparatively little research, however, has been done on the perception of polyphonic pitch (or harmony) in CI subjects. One recent study (Galvin *et al.*, 2009) examined melodic contour segregation in CI subjects using acoustically presented stimuli and found that CI users have difficulty segregating competing melodic contours even in the presence of timbral cues.

The objectives of the present study were to evaluate the ability of post-lingually deafened adult CI users to perceive the number of pitches in acoustically presented stimuli and to compare their performance with that of NH adults. Subjects listened to acoustically presented stimuli consisting of one, two, or three simultaneous tones with different fundamental frequencies (f_0) within a single octave. Both pure tones and piano tones were used to assess the effect of harmonics on polyphonic pitch perception. We hypothesized that CI users, as a result of diminished pitch resolution, would show decreased ability to differentiate between single versus mul-

Table 1. CI subject demographics.

Subject	Sex	Age	CI experience (years)	Device	Processor
CI1	M	69	5	ABC HiRes 90K	Harmony
CI2	F	58	6	CC Nucleus 24	Esprit 3G
CI3	M	27	2	ABC HiRes 90K	Harmony
CI4	F	68	1	CC Nucleus Contour	Freedom
CI5	M	58	5	ABC HiRes 90K	Harmony
CI6	M	46	2	CC Nucleus Contour	Freedom
CI7	F	33	4	CC Nucleus 24	Freedom
CI8	F	54	2	CC Nucleus Contour	Freedom
CI9	M	47	11	ABC Clarion	Platinum BTE
CI10	F	56	2	CC Nucleus Contour	Freedom
CI11	F	67	4	CC Nucleus Contour	Freedom
CI12	F	54	1	ABC HiRes 90K	Harmony

ABC: Advanced Bionics Corporation; CC: Cochlear Corporation; BTE: behind-the-ear.

tiple tones in comparison to NH controls. We further hypothesized that the ability of CI users to detect polyphony would increase as a function of interval distance between pitches, due to the presumptive relationship between increased frequency separation and improved perception of polyphony.

2. Methods

Twelve monaurally implanted CI users aged 27–69 years (mean=53.1±12.6 years) and 12 NH controls participated in the study. CI subjects used a variety of devices and processing strategies (Table 1). Each CI user had at least 1 year of experience using their implant system. The 12 NH adults ranged in age from 21 to 45 years (mean=28.4±9.2 years). All subjects completed a musical experience questionnaire to ascertain the extent of their musical training. No CI or NH subjects had musical training beyond an amateur level. All experiments were performed at the Sound and Music Perception Laboratory of Johns Hopkins Hospital, under an IRB approved research protocol. Informed consent was obtained for all participants.

All piano stimuli were recorded using Ivory Grand Piano Virtual Instrument (Synthogy) on the Apple LOGIC PRO 7.0 platform. Pure tones were generated using AUDACITY 1.2.5 (Dominic Mazzoni, open source). All stimuli were exactly 2.5 s in duration and were normalized by root-mean square power with equal-loudness contour adjustment using ADOBE AUDITION 3.0. Pure tone stimuli were given linear rise/decay ramps of 200 ms to reduce onset clicks. All stimuli consisted of pitches from within a central octave ranging in f_0 from 261 (C4) to 523 Hz (C5). Single-pitch stimuli consisted of either pure tones or piano tones from C4 to B4 (12 unique pitches and 24 total stimuli). Two-pitch stimuli consisted of either pure tone or piano tone representations of all 12 possible intervals within the range C4–C5 (1–12 semitone interval distance and 24 total stimuli). Three-pitch stimuli consisted of either pure tone or piano tone representations of six unique symmetric chords (equal interval spacing between lower/middle and middle/higher pitches) within the range C4–C5. No stimuli contained both pure and piano tones. For each three-pitch stimulus, interval spacing ranged from one to six semitones (for both lower/middle and middle/upper pitches). The pitches were equally spaced in order to maintain a consistent musical dispersion of pitches within each of the three-pitch stimuli and minimize the effect of varying intervals on perception of polyphony. Because only six unique arrangements of three equally spaced pitches are possible within a single octave, two sets of three-pitch stimuli were created (24 total stimuli for both piano and pure tones). Pitches used in the sets of two- and three-pitch stimuli were mathematically distributed symmetrically across the octave

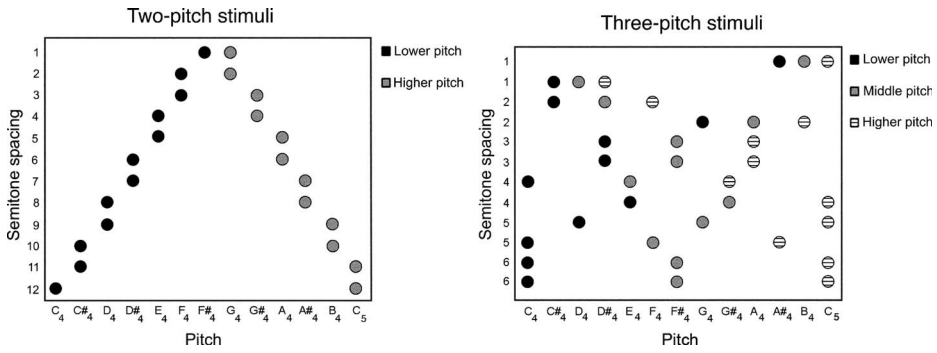


Fig. 1. Pitches and semitone spacing of two-pitch stimuli (left) and three-pitch stimuli (right).

so as to minimize over- or under-representation of any given pitch, as shown in Fig. 1. In total, 72 stimuli were presented to each subject.

Stimuli were randomly presented in a soundproof booth through a single calibrated loudspeaker (Sony SS-MB150H) at a presentation level of 80 dB sound pressure level through an OB822 clinical audiometer (Madsen Electronics); the speaker was positioned directly in front of the listener. For CI users, the contralateral ear (which was profoundly impaired in all individuals) was occluded with an earplug to diminish the effects of any minimal residual hearing, and no hearing aids were used. No subjects reported being able to hear any stimuli through their non-implanted ear. Stimuli were presented in a three alternative single-interval forced-choice procedure in which the subjects were instructed to choose whether the given stimuli consisted of one, two, or three pitches. Subjects were familiarized with the stimuli and procedure prior to formal testing. No feedback was given regarding the correctness of responses. The number of correct responses for each subject was averaged across the separate tone and pitch-number conditions to obtain an overall mean score.

3. Results

For all conditions, the CI group scored significantly lower than the NH group. The overall mean scores for each subject group were $43.1 \pm 12.3\%$ for CI users and $66.9 \pm 9.4\%$ for NH subjects. An unpaired t-test revealed a significant difference in overall mean scores between the NH and CI groups ($p < 0.001$). No statistically significant difference was found between average scores for pure tones and piano tones across subject groups. As none of the subjects had significant music training, musical experience was not analyzed as a covariate in this study. The mean scores of both subject groups for all stimuli, pure tones, and piano tones are shown in Fig. 2.

The CI group was significantly impaired in perceiving two- and three-pitch stimuli and scored close to 33% (i.e., near chance levels) when identifying two- and three-pitch stimuli (CI:

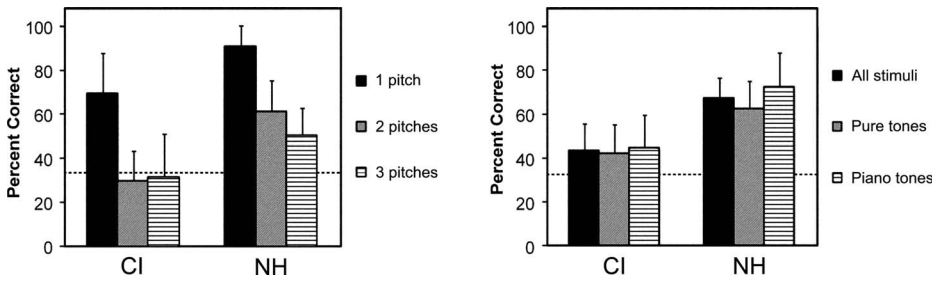


Fig. 2. Mean performance accuracy across CI subjects and NH subjects for one-, two-, and three-pitch stimuli (left) and for all stimuli, pure tones, and piano tones (right). The error bars show one standard deviation of the mean, and the dashed line shows chance performance level (33.3% correct).

Table 2. Confusion matrix for NH subjects.

		Identified			
		One pitch	Two pitches	Three pitches	No response
Presented	One pitch	260	24	2	2
		90.3%	8.3%	0.7%	0.7%
	Two pitches	63	172	53	0
		21.9%	59.7%	18.4%	0%
	Three pitches	23	122	143	0
		8.0%	42.4%	49.6%	0%

one-pitch – 69.1±18.6%, two-pitch – 29.1±14.1%, and three-pitch – 30.9±20.0%). In comparison, the NH group was much more successful at distinguishing single from multiple pitches, but demonstrated difficulties at distinguishing between two- and three-pitch stimuli (NH: one-pitch – 90.6±9.9%, two-pitch – 60.4±14.8%, and three-pitch – 49.6±13.0%). With the exception of 1 CI subject who achieved a higher mean score (72.2±11.5%) than 7 out of the 12 NH subjects, performance ranges for the 2 groups had little overlap. Unpaired *t*-tests revealed significant differences in scores between the NH and CI groups for all pitch conditions (one-pitch – $p=0.0026$, two-pitch – $p<0.001$, and three-pitch – $p=0.0136$). While NH subjects often identified three-pitch stimuli as having two pitches (suggesting that three-pitch condition was the most difficult for NH subjects), CI subjects often identified both three-pitch and two-pitch stimuli as a single pitch. NH subjects were less likely to identify two- and three-pitch stimuli as one pitch compared to CI users ($p=0.004$ for two-pitch stimuli and $p<0.001$ for three-pitch stimuli, unpaired *t*-test). Confusion matrices for NH and CI subjects are presented in Tables 2 and 3, respectively. For 4 out of 1728 total stimulus presentations (0.023%), the response period eventually timed out without a subject response. While this altered chance levels to a very small extent, we did not feel that this was a relevant variable to include in the analysis given the small magnitude of this effect.

Figure 3 shows the two-pitch performance accuracy of both subject groups as a function of interval distance. CI subjects performed near chance levels for all three-pitch conditions and most two-pitch conditions. For three-pitch conditions, there was no apparent relationship between interval spacing and ability to detect polyphony. For two-pitch conditions, increased interval spacing did not lead to better performance for detection of polyphony. In fact, an inverse relationship was suggested for identification of the one semitone interval spacing in two-pitch conditions (minor second interval), for which CI users were nearly as accurate as NH subjects.

Table 3. Confusion matrix for CI subjects.

		Identified			
		One pitch	Two pitches	Three pitches	No response
Presented	One pitch	199	63	26	0
		69.1%	21.9%	9.0%	0%
	Two pitches	136	84	67	1
		47.2%	29.2%	23.3%	0.3%
	Three pitches	112	86	89	1
		38.9%	29.9%	30.9%	0.3%

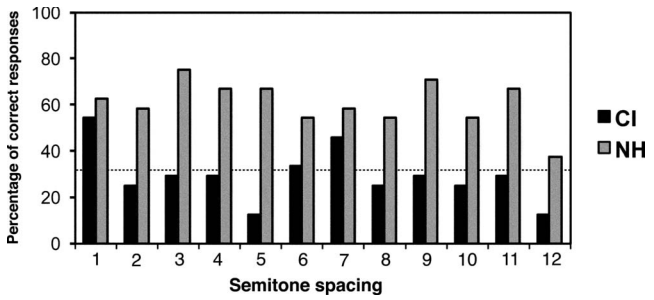


Fig. 3. Percentage of correct responses (number of correct responses/number of stimuli presented) across CI subjects and NH subjects as a function of semitone spacing in two-pitch stimuli. The dashed line shows chance performance level (33.3% correct).

4. Discussion

While previous studies have examined pitch resolution in CI subjects using pitch ranking and pitch discrimination tasks, the present study instead examined the ability of CI users to perceive acoustic polyphonic pitch using a novel pitch separation task in which subjects were asked to distinguish between one-, two-, and three-pitch stimuli. These categories of musical stimuli were chosen for their ubiquitous usage in Western music for melodies, intervals, and chords. The results from this study show that CI users obtain significantly lower average scores than NH subjects when asked to distinguish between single and multiple acoustically presented tones. CI users identified two- and three-pitch stimuli near chance levels, and demonstrated frequent perceptual fusion of multiple-pitch stimuli as single-pitch units. Both groups demonstrated a bias toward identifying all stimuli as one pitch (NH: one pitch—40%, two pitches—37%, and three pitches—23%; CI: one pitch—52%, two pitches—27%, and three pitches—21%). However, NH subjects were less likely to identify two- and three-pitch stimuli as one pitch. While a listener's ability to identify the number of components in a polyphonic stimulus does not necessarily correspond to the ability to perceive differences between polyphonic stimuli in a musical context (e.g., whether a musical triad is major or minor), perceptual fusion of polyphonic pitch likely impairs CI users in accurately perceiving many aspects of music, such as harmony, consonance, dissonance, and tonality.

One explanation for the lower average scores of CI subjects is the limited pitch resolution afforded by current CI devices. While pitch resolution varies widely across CI users, studies have shown that CI users rarely exhibit pitch discrimination thresholds of less than several semitones when two tones are presented sequentially (Gfeller *et al.*, 2002; Pretorius and Hanekom, 2008). Additionally, studies in NH listeners have shown that much greater frequency differences are required for the resolution of two tones sounding simultaneously than for the discrimination of two tones presented sequentially. For NH listeners, the limit for separation of two superposed pure tones within 100–500 Hz is roughly one semitone, almost ten times larger than the just noticeable difference for single pure tones within the same frequency range (Roederer, 1995). CI users would be expected to demonstrate even poorer separation of superposed tones than NH subjects due to poor pitch resolution.

CI subjects were most accurate in identifying two-pitch stimuli for both piano and pure tones when the two pitches were one semitone apart. This result was unexpected, given that sounds with more similar f_0 are more likely to be heard as a single stream (Oxenham, 2008). In CI subjects, all f_0's were likely processed within a single analysis band. Corresponding harmonics were also likely processed in the same bands. It is possible that temporal cues in the envelope extraction from the constructive and destructive interferences between two tones at slightly different frequencies aided CI subjects in the correct identification of the number of pitches present. This interference (also known as a beat) is perceived as periodic variations in volume whose rate is the difference between the two frequencies (Δf). Numerous studies have shown

that macroscopic temporal cues are readily perceived by CI users (Leal *et al.*, 2003), and it is possible that temporal envelope cues from beats aided CI subjects in separating superposed pitches. It is possible that beats provide optimal temporal cues to CI users when Δf is near one semitone, while temporal envelope cues with higher rates (when Δf is greater than one semitone) are less useful for helping CI users in separating polyphonic pitch.

The similarity in average scores between pure tones and piano tones across both subject groups was also an unexpected result. In the present study, it was anticipated that both subject groups would utilize the additional pitch information present in the overtones of the wider bandwidth piano tones to aid in the identification of polyphonic pitch. However, our results show that the presence of additional harmonic information does not necessarily increase the ability of both normal and CI subjects to perceive polyphony. Further research incorporating a larger set of complex tones and an examination of electrical stimulation patterns may be needed to better assess the effect of harmonics on pitch resolution in CI subjects.

5. Conclusion

CI users were found to obtain significantly lower scores than NH subjects when asked to distinguish between stimuli consisting of one, two, or three superposed tones, demonstrating perceptual fusion of multiple tones as single-pitch units. However, CI users were nearly as accurate as NH subjects at identifying the number of pitches present when two superposed tones were separated by one semitone. Overall, no statistically significant difference was found between average scores for pure tones and piano tones across both subject groups. This finding indicates that the presence of additional pitch information in complex tones may not aid either subject group in the resolution of polyphonic pitch. Perceptual fusion of polyphonic pitch likely contributes to the poor perception of harmony in CI users. As most of the music that CI users encounter is polyphonic, these findings indicate the need for further research on how polyphonic pitch is perceived by CI users. The development of processing strategies directed toward a more accurate representation of polyphonic pitch should greatly improve the ability of CI users to perceive complex musical stimuli.

Acknowledgment

The authors would like to thank all the subjects who participated in this study.

References and links

Galvin, J. J., Fu, Q., and Oba, S. I. (**2009**). "Effect of a competing instrument on melodic contour identification by cochlear implant users," J. Acoust. Soc. Am. **125**, EL98–EL103.
Gfeller, K., Turner, C., Mehr, M., Woodworth, G., Fearn, R., Knutson, J., Witt, S., and Stordahl, J. (**2002**). "Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults," Coch. Imp. Inter. **3**, 29–53.
Leal, M. C., Shin, Y. J., and Laborde, M.-I. (**2003**). "Music perception in adult cochlear implant recipients," Acta Oto-Laryngol. **123**, 826–835.
Looi, V., McDermott, H. J., McKay, C. M., and Hickson, L. (**2004**). "Pitch discrimination and melody recognition by cochlear implant users," Int. Congr. Ser. **1273**, 197–200.
McDermott, H. J. (**2004**). "Music perception with cochlear implants: A review," Trends Amplif. **8**, 49–82.
Oxenham, A. J. (**2008**). "Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants," Trends Amplif. **12**, 316–31.
Pijl, S. (**1997**). "Labeling of musical interval size by cochlear implant patients and normal hearing subjects," Ear Hear. **18**, 364–72.
Pretorius, L. L., and Hanekom, J. J. (**2008**). "Free field frequency discrimination abilities of cochlear implant users," Hear. Res. **244**, 77–84.
Roederer, J. G. (**1995**). "Superposition of pure tones: First order beats and the critical band," in *The Physics and Psychophysics of Music* (Springer-Verlag, New York), pp. 28–36.

Development of perceptual sensitivity to extrinsic vowel duration in infants learning American English

Eon-Suk Ko
University at Buffalo, State University of New York, Buffalo, New York 14260
eonsukko@buffalo.edu

Melanie Soderstrom
University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada
m_soderstrom@umanitoba.ca

James Morgan
Brown University, Providence, Rhode Island 02912
james_morgan@brown.edu

Abstract: 8- and 14-month-old infants' perceptual sensitivity to vowel duration conditioned by post-vocalic consonantal voicing was examined. Half the infants heard CVC stimuli with short vowels, and half heard stimuli with long vowels. In both groups, stimuli with voiced and voiceless final consonants were compared. Older infants showed significant sensitivity to mismatching vowel duration and consonant voicing in the short condition but not the long condition; younger infants were not sensitive to such mismatching in either condition. The results suggest that infants' sensitivity to extrinsic vowel duration begins to develop between 8 and 14 months.

PACS numbers: 43.71.Ft, 43.70.Mn, 43.71.Es [JH]
Date Received: July 14, 2009 **Date Accepted:** August 31, 2009

1. Introduction

The development of phonological category knowledge involves two important components. Infants' perceptual systems must be tuned to the phoneme boundaries that exist in their native language, and they must be sensitive to systematic subphonemic variations in which the location of phoneme boundaries is influenced by variations along other acoustic dimensions. One example of the latter type is the relationship between the vowel length and the perception of a coda consonant as voiced or voiceless. In this study, we examine infants' sensitivity to the property of consonant voicing in the context of short and long vowel durations.

Young infants demonstrate sensitivity to within-category subphonemic distinctions in voice onset time (Miller and Eimas, 1996). Similarly, infants are sensitive to the allophonic variation of aspiration in isolation at 2 months (Hohne and Juscyk, 1994) and are able to use allophonic information as a cue to identify familiarized target words in fluent speech by the age of 10.5 months (Jusczyk *et al.*, 1999). Infants are therefore able to detect at least some subphonemic variations, when they do not affect the perception of phoneme boundaries. What is unknown is how these sensitivities influence infants' phonological representations, when those variations are relevant to native-language-like perception of phoneme distinctions.

The present article investigates infants' development of perceptual sensitivity to subsegmental phonotactics, focusing on variation in vowel duration conditioned by the voicing of the following consonant. Vowels are realized with longer duration before a voiced than a voiceless consonant in English, e.g., [pɪk] vs [pɪːg] (House and Fairbanks, 1953). This effect will be referred to as "vowel length effect" (VLE). The duration of a pre-consonantal vowel thus serves as a source of information about the voicing of the following consonant. In addition to the VLE, earlier research has examined aspects of other cues for the post-vocalic voicing such as F1

offset frequency (Fischer and Ohde, 1990), intensity decay time, and the presence or absence of a "voice bar" during the closure interval (Hillenbrand *et al.*, 1984). The focus of our investigation was on the development of sensitivity to VLE-induced phonotactics.

Adult English listeners weight vocalic duration strongly in their perceptual decisions about the voicing of final stops, especially when no release burst is present (Denes, 1955) or the stimuli are synthetic (Raphael, 1972). However, 5–10 year old children and adults tested with stimuli based on natural utterances attend largely to dynamic signal components such as the F1-offset transitions rather than the vocalic duration (Morrongiello *et al.*, 1984, although see Hillenbrand *et al.*, 1984). Eilers (1977) suggested that infants at around 2 months of age use vowel duration as a supplementary cue for discriminating final consonantal voicing. Eilers *et al.* (1984) similarly found that infants (5–11 months) have the ability to discriminate vowel duration differences but their performance was much poorer than that of adults. Both studies examined instances of lengthening but not shortening. Lengthening differs from shortening, in that higher-level prosodic effects can also cause vowels to be lengthened, for example when words are focally or emphatically stressed, or occur at the ends of phonological or intonational phrases. These other factors may well complicate infants' reactions to vowel lengthening.

Recently, Dietrich *et al.* (2007) found that Dutch and English learning 18-month-olds treat vowel duration differently in a word learning task. In Dutch, vowel duration is an important cue for differentiating the low vowels [ɑ] and [aː], whereas in English, it is only a secondary cue to distinguish a tense from a lax vowel. Their results indicate that these properties are reflected in infants' perceptual sensitivity: Dutch learners interpret vowel duration as lexically contrastive, whereas English learners do not. One might therefore predict that 18-month-old English learners do not discriminate vowel duration differences. However, a subsequent study by Mugitani *et al.* (2009) found that 18-month-old English learners discriminate vowel duration differences if the task does not require linking objects with words. They also found that, in Japanese, where vowel length is phonemic, younger infants (10-month-olds) discriminate vowel duration differences like English 18-month-olds, while 18-month-olds show an asymmetric pattern of discrimination, responding to shortening, but not lengthening, of the vowel.

What do these findings suggest about infants' knowledge of the phonotactic patterns characterized by VLE? As noted, subphonemic variation in vowel duration can serve as a cue to post-vocalic voicing. Is infants' sensitivity to this pattern an innate characteristic of the perceptual system, or does it develop through exposure to the distributional characteristics of the language? Cross-linguistic comparisons suggest that speakers of languages without the VLE do not rely on vowel duration as a cue to voicing as much as do speakers of languages with the VLE. The use of the VLE as a perceptual cue may be learned through the experience with a native language (Crowther and Mann, 1992). Such language-specific patterns in perceptual weighting strategies predict that infants learning American English must acquire their sensitivity to the VLE at some point. Given that infants' speech perception is largely native-like by around 12 months (Werker and Tees, 1984), a year's exposure to English may have provided enough information for infants to develop their perceptual sensitivity to the VLE. However, there is relatively little work on their perception of coda consonants.

The present study investigated the development of 8- to 14-month-olds' perceptual sensitivity to the VLE. These ages roughly correspond to the beginning and end of the period of attunement toward native-like phoneme perception. Infants' first words also emerge toward the end of this period, giving us an opportunity to relate the results of their perceptual sensitivity to the patterns of the VLE in their early speech production. Recent findings (Ko, 2007) suggest that infants' learning of the VLE may have already begun to develop by the onset of their speech production. We hypothesized that infants by 14 months may have begun to develop their sensitivity to the VLE. We presented half the infants with CVC syllables containing a long vowel followed by a voiced (matched) or a voiceless (mismatched) consonant, and the other half with syllables containing a short vowel followed by a voiced or a voiceless consonant. If infants detect the relationship between vowel duration and coda voicing, they should discriminate matched from mismatched trials.

Table 1. Mean naturalness scores for mismatch tokens.

Token	bæːk	bæg	kʌːp	kʌb	pɪːk	pɪg
Mean naturalness score (n=3)	3.3	3.2	3.6	3.5	3.2	3.7

2. Method

Subjects. Seventy infants were tested; thirty-three 8-month-olds, and thirty-seven 14-month-olds. Four participants in the 14-month-old group were excluded from analysis because of fussiness (n=2) or lack of interest in the study (n=2). One participant in each of the 8- and 14-month-old groups was excluded due to experimenter error. This left thirty-two 8-month-olds (16 boys and 16 girls, mean age = 257 days, age range = 241–290 days) and thirty-two 14-month-olds (19 males and 13 females, mean age = 432 days, age range = 411–451 days). Half the infants heard words with a long vowel, followed by either a voiced (*matched long*; [pɪːg, kʌːb, bæːg]) or a voiceless consonant (*mismatched long*; [pɪːk, kʌːp, bæːk]). The other half heard *matched short* ([pɪk, kʌp, bæk]) and *mismatched short* ([pɪg, kʌb, bæg]) stimuli.

Stimuli. The stimuli were constructed from three minimal pairs ending in a voiced/voiceless plosive, *bag/back*, *cub/cup*, and *pig/pick*. A female native speaker of American English spoke the base words multiple times with infant-directed prosody, using a strong coda release. This ensured that perceptual cues for voice distinction associated with the release of a plosive were available in the stimuli, eliminating the possibility of cues other than vowel duration interfering as a confounding factor in the perception of the VLE-induced patterns.

Six exemplars of each word were chosen as the base tokens to produce the final stimuli by manipulating the duration of the vowel. The stimuli underwent lengthening/shortening of the vowel using the PSOLA resynthesis method available in PRAAT (Boersma and Weenink, 2007). Mismatched stimuli were constructed by lengthening or shortening the nucleus vowel of the base token, and matched stimuli were generated by lengthening or shortening the mismatched stimuli back to the original vowel duration. We generated the matched tokens through manipulation rather than using the natural base tokens to prevent any confounding effects of infants' perception of or preferences for natural vs manipulated stimuli. The resulting stimuli contain all the cues for the post-vocalic voice distinction such as pitch and formant transitions except for the vowel duration. The degrees of lengthening and shortening were 160% and 50% of the nucleus vowel in the base token.

The resulting 36 mismatched tokens (6 exemplars × 6 words) were rated for naturalness by ten adult subjects. The purpose of this testing was to ensure that the lengthened and shortened mismatched stimuli maintain about the same level of naturalness. The stimuli were presented in randomized order using PRAAT, and subjects scored the naturalness of each token from the scale of 1 (least natural) to 5 (most natural). Based on the results of the naturalness ratings, we selected 18 final tokens of mismatched stimuli (3 exemplars × 6 words) that yielded balanced naturalness ratings between lengthened and shortened tokens (see Table 1). The average vowel duration in the base tokens for these final tokens are reported in Table 2. Based on these 18 mismatched stimuli, we constructed 18 matched stimuli by manipulating the vowel duration back to the original base tokens.

Procedure. Testing was performed in a sound-attenuated room using the Headturn Preference Procedure. The testing booth consisted of a three-walled enclosure made of white

Table 2. Mean duration of the vowel in base tokens.

Token	bæːg	bæk	kʌːb	kʌp	pɪːg	pɪk
Mean duration in ms (n=3)	297.4	119.5	166.8	95.6	215.2	110.4

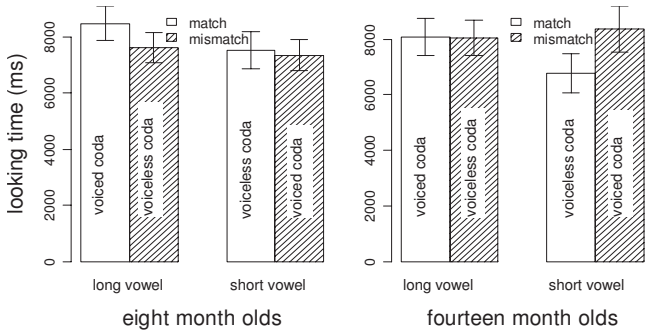


Fig. 1. Infants' preferences for VLE-matching vs mismatching word tokens.

pegboard panels, with a light mounted at the center of each panel wall. Caregivers sat with their infant on their lap and wore aviator headphones which played masking music to avoid biasing the infant's behavior. The order of trial presentation was randomized on-line by the experimental software.

Each trial began with the front light blinking to attract the infants' attention. When the infant looked at the center light, one of the two side lights began to flash. When the infant looked toward that light, the stimuli for that trial played from a speaker behind the light. Infants were first presented with two practice trials containing repetitions of three tokens of *book* and *dog*. They were immediately followed by a testing session of two randomized blocks of six trials containing the matched and mismatched versions of the 6 test words (12 test trials). Each trial consisted of random repetitions of the three exemplars of a particular word. Thus the "long" group heard tokens of [bæːg], [bæːk], [kʌːb], [kʌːp], [pɪːg], and [pɪːk] on successive trials, and the "short" group heard the short counterpart of each of these stimuli. Since word tokens varied considerably in length, pause durations between tokens for each trial were chosen in order to maintain a consistent interval between the onsets of each stimulus at 1200 ms. Therefore, infants heard similar rates of token presentation across trials and conditions. The dependent variable was the average amount of time each infant listened to matched vs the mismatched stimuli, based on their looking behavior.

3. Results

Mean looking times for stimuli with long and short vowels before voiced and voiceless coda consonants are shown in Fig. 1 for each of the two age groups tested. An analysis of variance (ANOVA) with two between-subjects factors, age and duration (short/long), and one within-subjects factor, matching, found significant interactions between age and matching, $F(1,60)=5.46$, $p<0.05$, and between duration and matching, $F(1,60)=4.36$, $p<0.05$ (see Fig. 1). Individual ANOVAs for each between-subjects condition found a significant interaction between matching and age, $F(1,30)=5.34$, $p<0.05$, and a marginal main effect of matching, $F(1,30)=3.47$, $p=0.072$, in the short condition, but no main effect or interactions in the long condition. In the 14-month-old age group, a marginal interaction was found for matching and duration, $F(1,30)=4.12$, $p=0.051$, with a marginal main effect of matching, $F(1,30)=3.79$, $p=0.061$. No significant effects were found with the 8-month-olds. Overall, these effects and interactions reflect a significant preference for the mismatched stimuli (mean listening time = 8.4 s) over the matched stimuli (mean listening time = 6.9 s) in the short condition for the older infants only, $t(15)=3.20$, $p<0.01$.

In sum, 14-month-olds showed a significant sensitivity to the mismatching of vowel duration and consonant voicing in the short, but not the long condition. Eight-month-olds did not show sensitivity with either short or long vowels.

4. Discussion

Our data suggest that sensitivity to the VLE develops over the course of the second half of the first year of life, consistent with the view that it is acquired through experience with phonotactic patterns in the native language. This is convergent with findings that speakers of languages without the VLE use vowel duration less than speakers of languages with the VLE (Crowther and Mann, 1992). It is also consistent with the recent finding that language-specific phonology influences the development of infants' speech perception (Mugitani *et al.*, 2009).

At first blush, our findings appear to contradict Dietrich *et al.* (2007), in which English-learning 18-month-olds failed to link two novel objects with the two stimuli differing only in vowel duration. However, there is good reason to suspect that older infants are less likely to discriminate auditory patterns in a word-learning context than in a pure preference or discrimination task (Stager and Werker, 1997). Therefore, it may be that 18-month-old English learners retain perceptual sensitivity to the VLE, as suggested in Mugitani *et al.* (2009), but fail to demonstrate this ability in a word-learning task: the oddness of a mismatch between vowel duration and coda voicing may not be regarded as encoding a lexical distinction.

The asymmetry between short and long vowels in our study may well be a consequence of infants' familiarity with vowel lengthening effects such as phrase-final lengthening and vowel elongation in infant-directed speech. Vowels are lengthened due to a variety of causes, and thus long vowels appear in variable contexts in the input. Therefore, infants may treat shortening as a more relevant cue for the phoneme boundary than lengthening or treat lengthening as more acceptable than shortening. This is consistent with our observation that 14-month-olds discriminated matched and mismatched exemplars containing short vowels, but not exemplars containing long vowels. Similar findings of such asymmetry are reported in other studies. For example, Hogan and Rozsypal (1980), testing the effects of vowel modulation on adults' judgment of voice distinction for post-vocalic consonants, reported findings of a pilot study in which recognition of the stimuli ending with a voiceless consonant remained unaffected by lengthening of the vowel. More recently, Japanese 18-month-old infants (Mugitani *et al.*, 2009) and Dutch 21-month-old toddlers (van der Feest and Swingley, 2008) have been reported to show asymmetric discrimination patterns to the vowel duration change. These findings suggest different processing of lengthening and shortening in infants as well as adults.

Given the stimuli we used, it is possible that 14-month-olds perceived short vowels preceding voiced consonants as aberrant pronunciations of familiar words, rather than as violations of more general phonotactic patterns. We plan to tease these possibilities apart in a follow-up study using nonce word stimuli.

Our results indicate that the perceptual system of 14-month-olds, who are at the beginning stages of word production, is already sensitive to the VLE, at least in some contexts. This suggests that the emergence of the VLE in children's early speech (Ko, 2007) may reflect children's knowledge of English phonotactics in the perceptual domain. The current study thus provides some concrete data to corroborate the idea that the development of speech production is preceded by the development of perceptual sensitivity.

5. Conclusion

The current study examined infants' development of perceptual sensitivity to the VLE. Our findings suggest that infants' sensitivity to the phonotactic patterns conditioned by the VLE begin to develop between 8 and 14 months. Infants may begin to use vowel duration as a cue to voicing at least as early as 14 months. Our results also point to an asymmetry supported by a growing body of research indicating that lengthening and shortening effects are treated differently in the speech perception.

Acknowledgments

This study was supported by NIH Grant No. R01 HD23005 to J.L.M. We thank Lori Rolfe, Elena Tenenbaum, Erin Conwell, Jae Yung Song, Amanda Seidl, Alex Cristià, and the participants of the experiments for their help in completing this study.

References and links

Boersma, P., and Weenink, D. (**2007**). PRAAT: Doing Phonetics by Computer Version 4.6.38, from http://www.praat.org/ (Last viewed November, 2009).

Crowther, C. S., and Mann, V. A. (**1992**). "Native language factors affecting use of vocalic cues to final consonant voicing in English," J. Acoust. Soc. Am. **92**, 711–722.

Denes, P. (**1955**). "Effect of duration on the perception of vocing," J. Acoust. Soc. Am. **27**, 761–764.

Dietrich, C., Swingley, D., and Werker, J. F. (**2007**). "Native language governs interpretation of salient speech sound differences at 18 months," Proc. Natl. Acad. Sci. U.S.A. **104**, 16027–16031.

Eilers, R. (**1977**). "Context-sensitive perception of naturally produced stop and fricative consonants by infants," J. Acoust. Soc. Am. **61**, 1321–1336.

Eilers, R., Bull, D., Oller, K., and Lewis, D. (**1984**). "The discrimination of vowel duration by infants," J. Acoust. Soc. Am. **75**, 1213–1218.

Fischer, R. M., and Ohde, R. N. (**1990**). "Spectral and duration properties of front vowels as cues to final stop-consonant voicing," J. Acoust. Soc. Am. **88**, 1250–1259.

Hillenbrand, J., Ingrisano, D. R., Smith, B. L., and Flege, J. E. (**1984**). "Perception of the voiced-voiceless contrast in syllable-final stops," J. Acoust. Soc. Am. **76**, 18–26.

Hogan, J., and Rozsypal, A. (**1980**). "Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant," J. Acoust. Soc. Am. **67**, 1764–1771.

Hohne, E., and Jusczyk, P. (**1994**). "Two-month-old infants' sensitivity to allophonic differences," Percept. Psychophys. **56**, 613–623.

House, A., and Fairbanks, G. (**1953**). "The influence of consonantal environment upon the secondary acoustical characteristics of vowels," J. Acoust. Soc. Am. **25**, 105–113.

Jusczyk, P., Hohne, E., and Bauman, A. (**1999**). "Infants' sensitivity to allophonic cues for word segmentation," Percept. Psychophys. **61**, 1465–1476.

Ko, E. (**2007**). "Acquisition of vowel duration in children speaking American English," in Proceedings of Interspeech 2007, pp. 1881–1884.

Miller, J., and Eimas, P. (**1996**). "Internal structure of voicing categories in early infancy," Percept. Psychophys. **58**, 1157–1167.

Morrongiello, B. A., Robson, R. C., Best, C. T., and Clifton, R. (**1984**). "Trading relations in the perception of speech by 5-year-old children," J. Exp. Child Psychol. **37**, 231–250.

Mugitani, R., Pons, F., Fais, L., Dietrich, C., Werker, J., and Amano, S. (**2009**). "Perception of vowel Length by Japanese- and English-learning infants," Dev. Psychol. **45**, 236–247.

Raphael, L. J. (**1972**). "Preceding vowel druation as a cue to the voicing characteristics of word-final consonants in English," J. Acoust. Soc. Am. **51**, 1296–1303.

Stager, C. L., and Werker, J. F. (**1997**). "Infants listen for more phonetic detail in speech perception than in word learning tasks," Nature (London) **388**, 381–382.

van der Feest, S., and Swingley, D. (**2008**). "A crosslinguistic study of vowel duration in 21-month-olds' early lexical representations," paper presented at the 16th International Conference on Infant Studies, Vancouver, Canada.

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

 [DOI: 10.1121/1.3244038] Published Online 8 October 2009

Straightforward estimation of the elastic constants of an isotropic cube excited by a single percussion

F. J. Nieves and F. Gascón
Departamento de Física Aplicada II, E.T.S. Arquitectura, Universidad de Sevilla, Reina Mercedes 2, 41012 Sevilla, Spain
nieves@us.es, fgascon@us.es

A. Bayón and F. Salazar
Departamento de Física Aplicada, E.T.S.I. Minas, Universidad Politécnica de Madrid, Ríos Rosas 21, 28003 Madrid, Spain
anaisabel.bayon@upm.es, felixjose.salazar@upm.es

Abstract: Ritz's method is applied to calculate accurate values of the lowest non-dimensional natural frequencies of a freely vibrating isotropic cube. The dependence of such frequencies and their quotients on Poisson's ratio is established. Vibration of a cube caused by percussion is detected at a point by a laser interferometer. With the help of the tables and graphs provided and with the values of the first lowest frequencies obtained experimentally in a single test, Poisson's ratio and the shear modulus are calculated by means of elementary arithmetical operations.

PACS numbers: 43.40.At, 43.40.Dx, 43.20.Ks [JM]
Date Received: April 30, 2009 **Date Accepted:** September 11, 2009

1. Introduction

As an antecedent of the work presented here, it is possible to mention the calculation of the dynamic elastic constants from the axisymmetric vibration of a cylinder whose length is equal to its diameter.[1] The measurement of two or three natural frequencies of vibration of the cylinder, originated by an axial percussion, enables the calculation of Poisson's ratio and the shear modulus.

A general analytical solution for the free vibration problem of a thick plate does not exist. An approximate solution can be obtained by Ritz's, finite element method (FEM), and other methods. In this paper the Ritz method is employed due to its advantages with respect to the FEM. Ritz's technique utilizes global basis functions, which are more accurate per degree of freedom than FEM. In addition, Ritz's method allows breaking the total vibration problem into smaller problems.[2] A solution for the amplitude of vibration of rectangular parallelepipeds was proposed in the form of power series of the coordinates.[2,3] Ritz's method has been optimized[4] in two-dimensional studies with the aim of simplifying calculations by means of an automatic search of the maximum exponents of the series, which gives a good convergence.

A three-dimensional calculation of the free vibration frequencies of an isotropic cube is here performed by the optimized Ritz method. It is demonstrated that the values of the five lowest natural frequencies enable the immediate calculation of the dynamic elastic constants.

2. Calculation of the natural frequencies of a free cube by the optimized Ritz method

For the sake of simplicity, the following non-dimensional frequency is used:

$$\Omega \equiv \pi f L \sqrt{\rho/G}, \tag{1}$$

where f is the ordinary frequency measured in hertz, L is the length of each side of the cube, G its shear modulus, and ρ its density.

Table 1. The five lowest, non-dimensional, different, non-null frequencies for a cube for six different values of Poisson's ratio and their modes.

Poisson's ratio	Ω_1	Ω_2	Ω_3	Ω_4	Ω_5
	1.427 418	1.866 441	1.936 756	2.169 072	2.221 442
0	*EV*1	*EX*1+*EX*2	*OX*1	*OX*2	*OD*1+*OD*2+*OD*3
	1.427 646	1.890 008	1.945 570	2.009 933	2.221 689
0.1	*EV*1	*EX*1	*OX*1	*EX*2	*OD*1
	1.427 879	1.908 425	1.951 381	2.151 210	2.221 445
0.2	*EV*1	*EX*1	*OX*1	*EX*2	*OD*1
	1.428 087	1.923 045	1.955 635	2.221 445	2.286 637
0.3	*EV*1	*EX*1	*OX*1	*OD*1	*EX*2
	1.428 234	1.934 846	1.958 959	2.221 445	2.412 044
0.4	*EV*1	*EX*1	*OX*1	*OD*1	*EX*2
	1.428 406	1.944 319	1.961 633	2.221 444	2.511 337
0.499	*EV*1	*EX*1	*OX*1	*OD*1	*OX*2

The Mindlin–Lamé modes[5] have an analytical solution; the lowest non-dimensional frequency is given by

$$\Omega_m = \pi/\sqrt{2}. \tag{2}$$

Let us suppose a harmonic solution for the displacements. Polynomials of monomials formed by products of powers of the coordinates are chosen for the amplitudes,[1] $U_i = \Sigma_{pqr} A_{ipqr} x_1^p x_2^q x_3^r$ ($i=1$, 2, and 3).

In applying Ritz's method, an optimization procedure[4] that improves the calculation is here generalized to three dimensions. The method involves initiating the calculation with a first stage in which very low maximum exponents are taken, and each exponent is then sequentially increased by one and the resulting frequencies are calculated and compared.

The calculation is based on the hypothesis that the first five non-null lowest frequencies are sufficient to determine the elastic constants of the cube. The reason why five frequencies are considered suitable was gathered from Fig. 1 of Ref. 6, in which it can be seen that some of the modes corresponding to those five frequencies have different behaviors with respect to Poisson's ratio. A criterion to decide which group of exponents is optimal in each stage could be the following: The best set is that for which the sum of the first five frequencies is minimum.

Using symmetry arguments of the displacement-field components it is concluded[7] that a single rectangular parallelepiped can vibrate in only eight basic forms or different groups of vibration. In the particular case of a cube, it can vibrate freely in the form defined by the four groups *OD* (dilatation), *EV* (torsion), *EX* (bending), and *OX* (shear), in agreement with the nomenclature of Heyliger *et al.*[8] Because of this reduction in modes, a cubic shaped sample is employed to make use of its high symmetry and simplify the calculation.

The values of Poisson's ratio used in the calculations are 0.0, 0.1, 0.2, 0.3, 0.4, and 0.499. This reduced number of values of ν, six, is chosen in order to avoid excessive calculations and extensive tables of results, but it allows interpolation, and therefore Poisson's ratio may be calculated.

The natural frequencies of the four indicated groups of vibration, numerically calculated and expressed to six decimal places, were shown in four basic tables. Table 1 summarizes the four basic tables.

The lowest natural frequency appearing in Table 1 corresponds to a torsion mode (mode *EV*1). Another identifiable frequency in the table is that of lower Mindlin, *OD*1, which, according to Eq. (2), is $\pi/\sqrt{2}$. In Table 1 none of its values differs from the theoretical value by more than 0.01%.

From the calculated frequencies, Fig. 1 is drawn, which represents the smallest Ω against ν for the lowest modes. Figure 1 has actually been drawn from each basic table by means

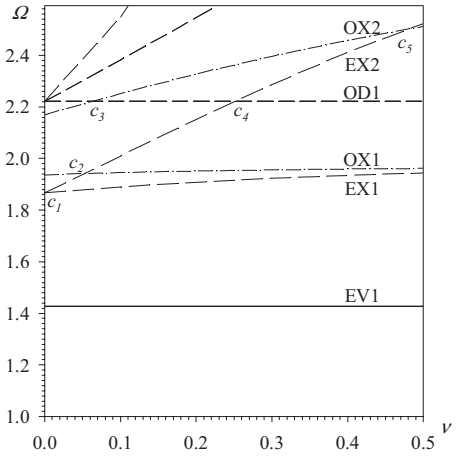


Fig. 1. The lowest computed non-dimensional frequencies Ω versus Poisson's ratio for a cube. The labels indicate the group of modes to which they belong.

of the parabolic fit of each of the columns Ω_i through the polynomial expression $\Omega_i = a_i + b_i\nu + c_i\nu^2$. In this figure it was observed that no curves corresponding to Ω superior to the sixth intersected with any of the six lower frequencies. None of the Ω frequencies is a decreasing function of ν. This figure is better, for low frequencies, than that published by Demarest,[6] at least for the reason that it includes values of Poisson's ratio from 0 to very near to 0.5.

Since Ω_{OD1} is independent of ν, and $\Omega_1 \equiv \Omega_{EV1}$ has such a small variation with ν, the ratio $\Omega_{OD1}/\Omega_1 \equiv \Omega_{OD1}/\Omega_{EV1}$ also has a small variation. In effect, from Table 1, the interval of possible values of $\Omega_{OD1}/\Omega_{EV1}$ is deduced to be from 1.5552 to 1.5563 and may be written as $\Omega_{OD1} = (1.5557 \pm 0.0006)\Omega_{EV1}$.

A similarly small interval results with the modes Ω_{OD1} and Ω_{OX1}, whose quotient is in the interval $\Omega_{OD1}/\Omega_{OX1} = 1.1397 \pm 0.0073$.

3. Calculation of the elastic constants of a cube

From the experimental natural frequencies, the quotients of all the pairs of the first five lowest frequencies f_i/f_j are calculated. From the definition of Ω, Eq. (1), it is deduced that the quotient $\Omega_i/\Omega_j = f_i/f_j$ only depends on ν. In Table 2, the numerical results for such quotients are listed.

Figure 2 shows the quotients of lowest frequencies versus ν and has been drawn from the parabolic fit and by applying the equalities

$$\frac{f_i}{f_j} \equiv \frac{\Omega_i}{\Omega_j} = \frac{a_i + b_i\nu + c_i\nu^2}{a_j + b_j\nu + c_j\nu^2}. \tag{3}$$

Table 2. The ten quotients between the five lowest frequencies for a cube, numerically calculated.

Poisson's	Ω_2/Ω_1	Ω_3/Ω_1	Ω_4/Ω_1	Ω_5/Ω_1	Ω_3/Ω_2	Ω_4/Ω_2	Ω_5/Ω_2	Ω_4/Ω_3	Ω_5/Ω_3	Ω_5/Ω_4
0	1.3076	1.3568	1.5196	1.5563	1.0377	1.1621	1.1902	1.1200	1.1470	1.0241
0.1	1.3239	1.3628	1.4079	1.5562	1.0294	1.0635	1.1755	1.0331	1.1419	1.1054
0.2	1.3365	1.3666	1.5066	1.5558	1.0225	1.1272	1.1640	1.1024	1.1384	1.0326
0.3	1.3466	1.3694	1.5555	1.6012	1.0169	1.1552	1.1891	1.1359	1.1693	1.0293
0.4	1.3547	1.3716	1.5554	1.6888	1.0125	1.1481	1.2466	1.1340	1.2313	1.0858
0.499	1.3612	1.3733	1.5552	1.7581	1.0089	1.1425	1.2916	1.1324	1.2802	1.1305

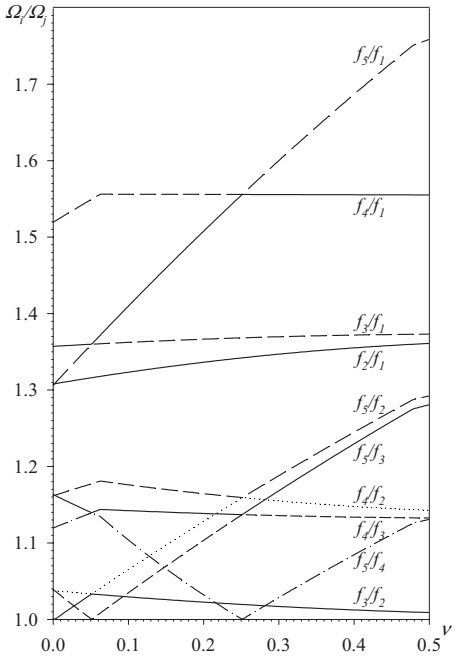


Fig. 2. The ten ratios Ω_i/Ω_j of the lowest frequencies versus Poisson's ratio for a cube.

The knowledge of the quotients obtained in the laboratory for the lowest frequencies f_i/f_j allows Poisson's ratio to be deduced from Table 2, or from Fig. 2, or from the equation of second degree in ν, Eq. (3). Observe that the calculation of ν is independent of the properties L, ρ, and G. However, it is necessary to be sure, by means of a detailed study, if ν is a single-valued function for some of the quotients of frequencies. The following should be borne in mind.

(a) Quotient Ω_3/Ω_2, for a typical value of $\nu=0.3$, is 1.0169, and therefore its relative difference is of the order of 2%. For this reason, frequencies f_3 and f_2 must be measured with high precision to be considered useful.
(b) Although the frequencies of Mindlin modes do not depend on ν, the quotients $\Omega_i/\Omega_{\text{Mindlin}} \equiv \Omega_i/\Omega_{OD}$ have the advantage that the denominator is known with great precision as is desired and all the quotients are increasing functions of ν.
(c) To ensure that the error in the estimation of Poisson's ratio is small, it is necessary for the absolute values of slopes of the curves in this Fig. 2 to be great for all the values of Poisson's ratio.
(d) Poisson's ratio ν is a single-valued function of the quotients Ω_i/Ω_j, only in certain intervals. Therefore, the knowledge of the first five lowest frequencies and their quotients does not assure, *a priori*, deduction of Poisson's ratio.

However, once the experimental frequencies are measured and ordered in increasing order, it is possible to determine Poisson's ratio by means of the following steps.

(1) Ten quotients f_i/f_j are calculated by always dividing the highest by the lowest, which guarantees quotients greater than the unit.
(2) Each of these quotients is marked on the vertical axis of Fig. 2. A horizontal straight line is drawn from each of these marks.
(3) The points where the horizontal straight lines cross the curves of Fig. 2, or at least the curves of greater slope, are marked, and, therefore, a set of crosses is obtained. Note that each horizontal line generally intersects the curves several times.

The vertical line drawn through ν of the cube crosses about ten times the curves of Fig. 2. Reciprocally, as a result of the measurement of the frequencies, and the accurate numerical results, ten crosses on the same vertical are obtained, and such a vertical line in turn intersects the horizontal axis at a point corresponding to the value of Poisson's ratio.

Since the measures of the frequencies are not perfect and neither are the calculated fit curves, the crosses are on a broken line around the vertical line through the expected value of ν. Therefore a graphical interpolation with a vertical straight line provides an average value for ν. Given Poisson's ratio, an experimental natural frequency f_i and its ith order spectrum, either Table 1 or Fig. 1, provide the value of the corresponding Ω_i. With this value of Ω_i, measured length L of the edge of the cube, its mass, and calculated density ρ, the application of the definition of Ω, Eq. (1), gives the value of the shear module.

A practical guide for the estimation of elastic constants can be constructed to avoid the main difficulties of the proposed method: the multiple-valued relation of ν with the quotient of experimental frequencies, and the high uncertainty that may appear in certain quotients. To this end the values of Poisson's ratio have been numerically calculated corresponding to the five intersections c_i that appear in Fig. 1: $\nu_{c1}=0.0020$, $\nu_{c2}=0.0512$, $\nu_{c3}=0.0626$, $\nu_{c4}=0.2511$, and $\nu_{c5}=0.4783$. The values of the quotients of frequencies, at the crossings that are of more interest here, are $\Omega_2/\Omega_1]_{\nu c1}=1.3080$, $\Omega_3/\Omega_1]_{\nu c2}=1.3599$, and $\Omega_4/\Omega_1]_{\nu c3}=1.3717$. Some rules leading to the identification of the interval of existence of ν are proposed: First, the lowest frequency always corresponds to the mode $EV1$, and the non-dimensional frequency Ω_1 may be assigned to f_1. Second, application of the ratio given in Sec. 2, $\Omega_{OD1}/\Omega_{EV1}=(1.5557\pm0.0006)$, may give the frequency f_{OD1} of the first Mindlin mode. Third, if, from the spectrum f_i, it is deduced that f_{OD1} is the fourth frequency, then Fig. 2 implies that the material under study has Poisson's coefficient $\nu\geq\nu_{c4}$. With this an interval of existence of ν is defined, which allows discernment to be made between assignable multiple values ν for a certain ratio f_i/f_j in Fig. 2. Fourth, if, for instance, it is deduced that f_{OD1} is the fifth frequency, then $\nu_{c3}\leq\nu<\nu_{c4}$. With this interval of existence, ν is determined. Note that between each pair of neighboring crossings of the curves in Fig. 2, ν is a single-valued function of the quotients.

The calculations of ν and of G may be made graphically by placing Fig. 1 on Fig. 2 so that their axes are aligned. In effect, knowledge of the point of crossing of a quotient Ω_i/Ω_j with a curve of Fig. 2 is enough to draw a descending vertical line to find ν. The same vertical, but in its ascending sense cuts the line of one of the modes drawn in Fig. 1 at a point from which a horizontal line leads to Ω_i and Eq. (1) gives G.

To increase accuracy an easy analytic procedure may be followed by applying Eq. (3) to an adequate quotient and by solving the second degree equation to obtain ν.

4. Experimental arrangement and experimental results

The cube under study is of commercial stainless steel, annealed, and of dimensions $a=50.05$ mm, $b=50.04$ mm, and $c=50.03$ mm: an almost perfect cube. The density of the cube is $\rho=7836$ kg/m^3. The elastic constants have been obtained from measurements of the P and S wave velocities in three perpendicular directions and with two polarizations. The values calculated for ν belong in the interval (0.2866, 0.2881), and their average value is $\nu=0.2873$. The values calculated for G are in the interval (77.55, 77.97), and their average value is $G=77.76$ GPa.

Measurement of the natural frequencies of vibration is made with a laser interferometer *I-O*.[1] The interferometer detects displacements in the order of magnitude of 1 nm, and the bandwidth of detection is of the order of tens of MHz. The fast Fourier transform (FFT) of the out-of-plane displacement component enables us to obtain the natural frequencies of vibration of the cube. The sample is supported on a rubber. The location of the rubber does not affect the frequencies measured. Advantages of this method with respect to resonant ultrasound spectroscopy (RUS) is that it is a non-contact technique, the vibration spectrum is obtained in one single test, and it does not need an iterative process to obtain the frequencies with which Poisson's ratio and the shear modulus are directly determined.

4.1 Impact and detection at two corners

In a first experiment one percussion is applied near to a corner and the consequent vibration is detected at the diagonally opposite corner. The five lowest frequencies are $f_1=28\ 625$ Hz, $f_2=38\ 450$ Hz, $f_3=39\ 125$ Hz, $f_4=44\ 475$ Hz, and $f_5=45\ 450$ Hz. If all the lowest frequencies have been detected, the lowest from among them must correspond to a torsion, i.e., $f_1=f_t$.

The quotients of the first five lowest frequencies obtained experimentally are $\Omega_2/\Omega_1=1.3430$, $\Omega_3/\Omega_1=1.3668$, $\Omega_4/\Omega_1=1.5537$, $\Omega_5/\Omega_1=1.5878$, $\Omega_3/\Omega_2=1.0176$, $\Omega_4/\Omega_2=1.1567$, $\Omega_5/\Omega_2=1.1821$, $\Omega_4/\Omega_3=1.1367$, $\Omega_5/\Omega_3=1.1617$, and $\Omega_5/\Omega_4=1.0219$. These quotients of frequencies can be compared with those of Table 2, be substituted in Eq. (3), and be marked in Fig. 2.

The quotient $\Omega_{OD1}/\Omega_{EV1}=1.5557\pm0.0006$, obtained previously by numerical methods, does not appear in the series of experimental quotients. The nearest is Ω_4/Ω_1 whose difference with respect to the theoretical value is 0.13%, which can be due to small errors in the measurement and geometry of the sample. However, the quotient $\Omega_{OD1}/\Omega_{OX1}=1.1397\pm0.0073$ appears among the experimental quotients (Ω_4/Ω_3). Therefore the fourth frequency of the spectrum corresponds to the mode $OD1$.

Let us see that the methodology to find ν is still correct and can be applied to the sample studied in spite of the small errors in the spectrum of frequencies. Since it has been deduced that f_{OD1} is the fourth frequency, the material under study has Poisson's ratio $\nu\geq\nu_{4cs}=0.2511$. This leads to the formation of an interval of existence for ν, which in turn enables the determination of the value of ν for a certain ratio f_i/f_j with Table 2, Fig. 2, or Eq. (3).

The quotient of greatest slope in Fig. 2 is f_5/f_1. For the experiment carried out $\Omega_5/\Omega_1=1.5878$; this quotient should be in the column of Ω_5/Ω_1 in Table 2 in the interval from lower value $(\Omega_5/\Omega_1)_l=1.5558$ and the higher value $(\Omega_5/\Omega_1)_h=1.6012$. By a linear interpolation, $\nu=0.2705$ is obtained. Another independent consequence is deduced: A simultaneous interpolation in Tables 1 and 2 gives directly Ω_k from the double linear interpolation formulas

$$\frac{\Omega_k-\Omega_{kl}}{\Omega_{kh}-\Omega_{kl}}=\frac{\Omega_i/\Omega_j-(\Omega_i/\Omega_j)_l}{(\Omega_i/\Omega_j)_h-(\Omega_i/\Omega_j)_l}, \tag{4}$$

and then G may be calculated from Ω_k and f_k. In this way Ω_1 or Ω_5 is found. For example, the table gives $\Omega_1=1.428\ 025$. Application of Eq. (1) with the pair Ω_1 and f_1 gives $G=77.81$ GPa, whose difference with the value calculated from the velocity measurements is 0.06%. Note that either of the elastic constants can be calculated independently of the other.

In the interval of existence of ν, the most suitable quotients due to their greatest slope are, in order of better to worse, f_5/f_1, f_5/f_4, f_5/f_3, and f_5/f_2.The application of Eq. (3) gives the respective values of Poisson's ratio 0.2863, 0.2884, 0.2896, and 0.2899. This set has an average value $\nu=0.2870$, whose difference with that calculated from the P and S wave velocities is 0.10%.

As in the zone of existence of ν, Fig. 1, the arrangement of the modes in order of increasing frequencies is $EV1$, $EX1$,…, the correspondence of f_i with Ω_i is known, and Eq. (1) may be applied. The resulting value of G for frequency f_4 (mode $OD1$) is 77.62 GPa.

4.2 Perpendicular impact and detection at the centers of opposite faces

For this kind of excitation and detection and by symmetry considerations, it is deduced that the detectable modes must be in the OD or EX mode groups. The detected frequencies are $f_1=44\ 475$ Hz, $f_2=45\ 450$ Hz,…. Quotient $f_2/f_1=1.0219$ may correspond to the quotient Ω_5/Ω_4 according to Fig. 2 and Table 2. Therefore the first three lowest frequencies have not been detected.

Application of Eq. (3) to this pair of detected frequencies gives $\nu=0.2884$ whose difference with the value obtained from the velocity measurements is 0.38%.

If the pair f_1, Ω_4 is used to calculate G, then application of Eq. (1) gives directly $G=77.62$ MPa, which leads to a difference of 0.18% with respect to the value calculated from the velocity measurements.

Observe that either of the elastic constants can be calculated independently of the other. The appearance in the spectrum of the Mindlin frequencies, whose Ω is independent of ν, and of the second frequency of bending ($EX2$) has facilitated an accurate and rapid calculation of the two elastic constants, as was foreseen.

4.3 Systematic uncertainty

The systematic uncertainty U of the measurement is estimated. Let us suppose resolution to be the only source of uncertainty. The frequencies are measured directly in the experiments carried out; its systematic uncertainty is $U_f=25$ Hz. For representative values in this paper: $\nu=0.3$, $f=40$ kHz, and the relative uncertainty of Poisson's ratio for the quotient $\Omega_{EX2}/\Omega_{OD1}$ is $U_\nu/\nu=\left|\partial\nu/\partial(\Omega_i/\Omega_j)\right|U_{\Omega_i/\Omega_j}/\nu\approx 4\%$. For the quotient $\Omega_{EX2}/\Omega_{OD1}$, $U_\nu/\nu=0.1\%$. Note the importance of an adequate quotient f_i/f_j to accurately calculate ν.

From Eq. (1) the relative systematic uncertainty of the shear modulus gives $U_G/G=2U_f/f+U_m/m+U_L/L+2U_\Omega/\Omega=12.5\times10^{-4}+3.48\times10^{-4}+2.00\times10^{-4}+0.01\times10^{-4}=0.2\%$, where $U_m=10^{-4}$ kg, $U_L=10^{-5}$ m, $U_\Omega=10^{-6}$, and $\Omega=2$ have been taken. The most important source of error of the shear modulus is the uncertainty of frequency.

References and links

[1] F. J. Nieves, F. Gascón, and A. Bayón, "Estimation of the elastic constants of a cylinder with a length equal to its diameter," J. Acoust. Soc. Am. **104**, 176–180 (1998).

[2] P. Heyliger, P. Ugander, and H. Ledbetter, "Anisotropic elastic constants: Measurement by impact resonance," J. Mater. Civ. Eng. **13**, 356–362 (2001).

[3] R. Holland, "Resonant properties of rectangular piezoelectric ceramic parallelepipeds," J. Acoust. Soc. Am. **43**, 988–997 (1968).

[4] F. J. Nieves, A. Bayón, and F. Gascón, "Optimization of the Ritz method to calculate axisymmetric natural vibration frequencies of cylinders," J. Sound Vib. **311**, 588–596 (2008).

[5] T. Lee, R. S. Lakes, and A. Lal, "Resonant ultrasound spectroscopy measurement of mechanical damping: Comparison with broadband viscoelastic spectroscopy," Rev. Sci. Instrum. **71**, 2855–2861 (2000).

[6] H. H. Demarest, "Cube-resonance method to determines the elastic constants of solids," J. Acoust. Soc. Am. **49**, 768–775 (1971).

[7] I. Ohno, "Free vibration of a rectangular parallelepiped crystal and its application to determination of elastic constants of orthorhombic crystals," J. Phys. Earth **24**, 355–379 (1976).

[8] P. R. Heyliger, H. Ledbetter, S. Kin, and I. Reimanis, "Elastic constants of layers in isotropic laminates," J. Acoust. Soc. Am. **114**, 2618–2625 (2003).

A Bayesian approach to modal decomposition in ocean acoustics

Zoi-Heleni Michalopoulou
Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, New Jersey 07102
michalop@njit.edu

Abstract: A Bayesian approach is developed for modal decomposition from time-frequency representations of broadband acoustic signals propagating in underwater media. The goal is to obtain accurate estimates and posterior probability distributions of modal frequencies arriving at a specific time and their corresponding amplitudes, which can be employed for geoacoustic inversion. The proposed approach, optimized via Gibbs sampling, provides uncertainty information on modal characteristics via the posterior distributions, typically unavailable from traditional methods.

PACS numbers: 43.60.Hj, 43.30.Pc, 43.60.Jn [JC]
Date Received: July 23, 2009 **Date Accepted:** September 9, 2009

1. Introduction

The evolution of the frequency content of an acoustic signal with time often acts as a fingerprint of the propagation medium, particularly with broadband signals with frequencies of a few hundreds of hertz propagating long distances in underwater environments.[1] Intra- and inter-modal dispersions reveal significant information on properties of the waveguide. Dispersion is captured in arrival time differences between distinct frequencies within the same mode or across a number of propagating modes. Estimates of such time differences and also modal amplitudes can be employed in conjunction with optimization for source localization and environmental parameter estimation;[2] dispersion estimation for inversion in underwater acoustics has been typically pursued with simple short time Fourier transforms (STFTs) and wavelet analysis.[3–5]

In Ref. 6, it was demonstrated that, although arrival time-frequency estimation is limited in terms of resolution when Fourier transforms are used, more reliable time-frequency pairs can be extracted via the stationary phase approximation. It was shown that the spectrum of one mode in the received acoustic signal calculated with a STFT is a squared, shifted, and scaled sinc function centered at the modal frequency arriving at that particular time. Accurately estimating the center of the sinc pulse with a fitting process provides a modal frequency-time pair that can be used for successful geoacoustic inversion with limited uncertainty.[6] Estimation of the scaling of the squared sinc provides information on the modal amplitude, linked to attenuation in the propagation medium. Extending this approach, we can formulate a superposition of similar functions to represent multiple modes arriving simultaneously. In Ref. 7, we followed a similar concept, empirically fitting a sum of Gaussian pulses to a STFT frame, forming the observation equation of a particle filter.

In this work we improve on the modal frequency and amplitude estimation of Zorych and Michalopoulou[7] by employing smoothed Wigner-Ville distributions (WVDs) and an approximation similar to the one proposed in Ref. 6. Working in a Bayesian framework for dispersion estimation,[8] we calculate posterior probability distributions of the unknown modal frequencies and amplitudes and optimize the process with a Gibbs sampler in a manner that is analogous to the time-delay estimation approach of Michalopoulou and Picarelli.[9] This paper is organized as follows: Sec. 2 briefly presents the sinc approximation for modal Wigner-Ville time-frequency representations. Section 3 discusses the statistical model employed in this work for spectral observations and describes the Bayesian modal decomposition approach using the introduced models and a Gibbs sampler. Section 4 presents results of frequency and amplitude estimation of distinct modes. A summary and conclusions follow in Sec. 5.

2. Approximating the acoustic signal in time-frequency

The WVD is selected for calculating the time-frequency content of acoustic signals because it does not have the resolution limitations of Fourier transforms and can accurately represent dispersion at any pair $(\omega,t)=(2\pi f,t)$.[10] Such a time-frequency representation of an analytic signal $s_{a,n}$ is obtained as[10] $WVD(\omega,t)=1/2\pi\int_{-\infty}^{\infty}s_{a,n}(t-\tau/2)^{*}s_{a,n}(t+\tau/2)\exp(-i\tau\omega)d\tau$, where * indicates conjugate. In our application, employing a normal mode approach to model sound propagation in the ocean, we receive acoustic signal $p(r,z,z_0,t)$, where

$$p(r,z,z_0,t)=\sum_n s_{a,n}(r,z,z_0,t)=\frac{1}{2\pi}\sum_n\int_{-\infty}^{\infty}\mu(\omega)G_n(r,z,z_0,\omega)\exp\left(i\left(\omega t-k_n r-\frac{\pi}{4}\right)\right)d\omega. \tag{1}$$

Quantity r is the distance between source and receiver, z and z_0 are the source and receiver depths, respectively, k_n is the modal wavenumber, μ is the source spectrum, $\omega=2\pi f$ is frequency, and $G_n(r,z,z_0,\omega)=(\beta/\sqrt{k_n r})\Psi_n(z)\Psi_n(z_0)$, where Ψ_n are orthogonal, depth-dependent functions that are appropriately normalized and β is a constant (the waveguide is assumed to have a constant density).

Since we have multiple modes in our received acoustic field, the Wigner-Ville representation will include in the integrand a signal that is formed by the superposition of several modal contributions $s_{a,n}$: $s_a=p(r,z,z_0,t)=\Sigma_n s_{a,n}$. The time-frequency distribution will then have numerous cross-factors that will make its interpretation very difficult.[10] To avoid this complication, we use a smoothed, pseudo-Wigner-Ville Distribution (SPWVD), applying windows of finite length to multimodal signals s_a, smoothing away cross-terms. Assuming a window of length $2\Delta\tau$ and employing the stationary phase approximation,[11,6] the SPWVD of mode n in our signal becomes

$$\mathrm{SPWVD}_n(\omega,t)\approx\alpha_n\frac{\sin(\omega-\omega_n)\Delta\tau}{\omega-\omega_n}, \tag{2}$$

where α_n contains all multiplicative constants. Optimal window lengths $2\Delta\tau$ depend on the actual modal dispersion and separation (they, here, are empirically selected); specific window types (Kaiser, Hamming, and Blackman, for example) can be incorporated in the SPWVD.

The SPWVD of an acoustic signal can then be written as a superposition of weighted and shifted sinc functions, with a_n containing all multiplicative constants:

$$\mathrm{SPWVD}(\omega,t)\approx\sum_n\alpha_n\,\mathrm{sinc}((\omega-\omega_n)\Delta\tau). \tag{3}$$

3. The statistical model

The conventional assumption is that the time domain acoustic signal is distorted by additive, complex, zero-mean Gaussian noise for each channel (real and imaginary). Frequency representations, involving squared signal components, are typically then related to χ^2 distributions. The integral engaged in the calculation of WVDs (and SPWVDs) implies a summation of several χ^2 random variables. Consequently, following Ref. 7, a Gaussian model is used to describe random fluctuations in our observations. Thus, the misfit between our calculated SPWVDs and our parametric model is Gaussian distributed with zero mean and a variance of σ^2.

Given the sinc expression, the M elements of data vectors $\mathbf{W}$ can be written as

$$W_i=\sum_{n=1}^{N}\alpha_n\,\mathrm{sinc}((\Omega_i-\omega_n)\Delta\tau)+U_i, \tag{4}$$

where $\mathbf{U}=(U_1,U_2,\ldots,U_M)$ is a white Gaussian noise vector with $\mathbf{0}$ mean and variance σ^2, N is

the number of considered modes, and M is the number of frequencies Ω_i at which we sample the SPWVD.

Once $\mathbf{W}$ is observed, Eq. (5) provides the likelihood function for unknown modal amplitudes, frequencies, and variance σ^2 [$\mathbf{\Omega}=(\Omega_1,\ldots,\Omega_M)$]:

$$l(\alpha_1,\ldots,\alpha_N,\omega_1,\ldots,\omega_N,\sigma^2|\mathbf{W})=\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^M \exp\left(-\frac{1}{2\sigma^2}\left\|\mathbf{W}-\sum_{n=1}^{N}\alpha_n \operatorname{sinc}((\mathbf{\Omega}-\omega_n)\Delta\tau)\right\|^2\right). \tag{5}$$

Working in a Bayesian framework,[12,13] after establishing a statistical model for the time-frequency representations [Eq. (5)], we formulate the posterior probability distribution of the unknown parameters α_n,ω_n, $n=1,\ldots,N$, and σ^2. Bayes theorem allows us to express the posterior distribution as $p(\alpha_1,\ldots,\alpha_N,\omega_1,\ldots,\omega_N,\sigma^2|\mathbf{W})=Kl(\alpha_1,\ldots,\alpha_N,\omega_1,\ldots,\omega_N,\sigma^2|\mathbf{W})\times p(\alpha_1,\ldots,\alpha_N,\omega_1,\ldots,\omega_N,\sigma^2)$, where $p(\alpha_1,\ldots,\alpha_N,\omega_1,\ldots,\omega_N,\sigma^2)$ is the joint prior distribution on the unknown parameters and K is a normalizing constant such that $p(\alpha_1,\ldots,\alpha_N,\omega_1,\ldots,\omega_N,\sigma^2|\mathbf{W})$ integrates to 1.

Imposing no *a priori* correlation structure between unknowns, the joint prior distribution can be written as $p(\alpha_1,\ldots,\alpha_N,\omega_1,\ldots,\omega_N,\sigma^2)=p(\alpha_1)\cdots p(\alpha_N)p(\omega_1)\cdots p(\omega_N)p(\sigma^2)$, where the factors are the prior distributions for the individual variables. We select uniform prior distributions within chosen intervals (positive for the amplitudes and between 400π and 1200π for ω_n, because of the frequency content of our synthetic signals). The prior distribution for σ^2 is $1/\sigma^2$, which is the conventional non-informative prior for variance.[14]

To make the implementation of the Bayesian processor practical in the presence of multiple modes, we build a Gibbs sampler[15,16] for the estimation of $p(\alpha_1,\ldots,\alpha_N,\omega_1,\ldots,\omega_N,\sigma^2|\mathbf{W})$. Gibbs sampling is a Markov chain Monte Carlo approach that estimates the full joint posterior distribution of all unknown parameters given a set of observations, a statistical model describing uncertainty in the data, and prior distributions on the unknowns. This posterior distribution can be maximized in a straightforward manner for the calculation of maximum *a posteriori* (MAP) estimates.

Gibbs sampling is an iterative approach that starts with a set of randomly selected initial values for the unknown parameters. Sequentially, samples are drawn from the posterior distributions of each parameter conditional on all other parameters. For our purposes, we need to identify the $2N+1$ conditional marginal posterior distributions for α_n, ω_n, $n=1,\ldots,N$, and σ^2.

Focusing on parameter α_1 as an example, its posterior distribution conditional on all other parameters is found to be a Gaussian distribution with variance σ^2 and mean $(\Sigma_{i=1}^{M}W_i \operatorname{sinc}((\Omega_i-\omega_1)\Delta\tau)-\Sigma_{j=2}^{N}\alpha_j \operatorname{sinc}((\Omega_i-\omega_1)\Delta\tau)\operatorname{sinc}((\Omega_i-\omega_j)\Delta\tau))$,[9] from which we can expeditiously draw samples. Conditional distributions for $\alpha_2,\ldots,\alpha_N$ are similarly obtained. The conditional distribution of σ^2 is simply expressible in a closed form as well and is identified as an inverse χ^2 distribution.[14,9] The distributions for ω_n cannot be expressed in a closed form and are evaluated on a grid.[9]

Gibbs sampling starts with a set of randomly chosen initial conditions for all unknown parameters $(\alpha_1,\alpha_2,\ldots,\alpha_N,\omega_1,\ldots,\omega_N,\sigma^2)$. The process as implemented here first draws a sample from the conditional distribution of α_1 given the initial values for all other parameters. Subsequently, a sample is drawn from the marginal conditional posterior of α_2 given initial values for the other parameters and the updated value for α_1, obtained during the previous step. The process is repeated for remaining α_n and ω_n and σ^2 as well. For a large number of iterations, the obtained sample sequence converges to the true joint posterior distribution. We can use this sequence to calculate marginal distributions and moments.

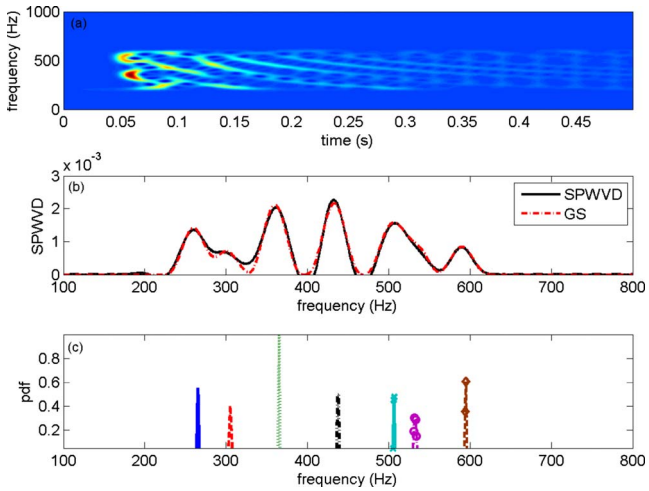


Fig. 1. (Color online) (a) The SPWVD of the acoustic signal. (b) A "slice" from the SPWVD representation of (a) (solid line); the best fit calculated via the Gibbs sampler is superimposed (dot-dashed line). (c) The marginal posterior probability distributions for $f_1, \ldots, f_7$.

4. Gibbs sampling results in modal decomposition

To first provide an example for the SPWVD of an acoustic signal, we have generated a synthetic acoustic reception in a shallow water environment similar to that for the Gulf of Mexico experiment.[17] The simulated source had a uniform, flat spectrum with frequency content between 200 and 600 Hz; the sampling rate was 2000 Hz, suitable for preventing aliasing in the WVD calculations for an analytic signal. The signal was calculated at a distance of 20 km from the source. Receiver and source depths and the water column sound speed profile are provided in Ref. 17; the seafloor sediment was a thin layer of sand over limestone. Figure 1(a) shows the SPWVD of the received signal, which was calculated with Kraken[18] and Fourier synthesis. A "slice" of the distribution at a selected time t is demonstrated in Fig. 1(b) (black, solid line).

The Gibbs sampler developed here was applied to the slice of the SPWVD (obtained by selecting a specific time t) of Fig. 1; the MAP estimate calculated with the method proposed in this work is superimposed in Fig. 1(b) (dot-dashed line) on the actual SPWVD representation at the selected time, indicating an excellent fit between the data and the parametric model involving sinc pulses. Figure 1(c) shows the marginal posterior probability distributions for seven modal frequencies.

The same technique was applied to a SPWVD of a signal with a low signal-to-noise ratio (SNR); Figure 2(a) demonstrates the MAP estimate of the SPWVD superimposed on the noisy SPWVD. The match is still excellent despite the noise level, demonstrating the potential of the proposed approach in parametrizing received acoustic time series in a manner that will enable geoacoustic inversion even for a low SNR. Figure 2(b) shows marginal posterior distributions for modal frequencies $f_1, \ldots, f_7$ as extracted from the noisy time-frequency representation. Figure 2(c) shows the joint posterior distribution for α_1 and α_2; amplitude information can be extracted in this manner to be subsequently employed in attenuation estimation.

Figure 3 illustrates real data results from application of the method to data from the Gulf of Mexico experiment (the frequency content was between 100 and 600 Hz). Figure 3(a) shows the SPWVD slice and the best fit. Data and parametric fit match very closely. Figure 3(b) presents the modal frequency distributions, indicating that the first "peak" in the selected SPWVD slice corresponds to two modes that are close to each other. Figure 3(c) shows the joint posterior distribution for α_3 and α_5. Although probability is mostly concentrated around 0.55

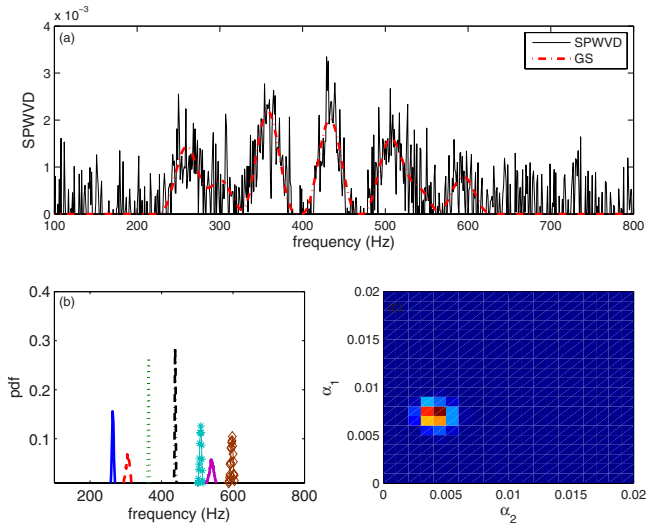


Fig. 2. (Color online) (a) A "slice" of a noisy SPWVD representation (solid line); the best fit calculated via the Gibbs sampler is superimposed (dot-dashed line). (b) The marginal posterior probability distributions for $f_1, \ldots, f_7$. (c) The joint marginal distribution for α_1 and α_2.

and 0.09 for the two amplitudes, secondary distribution peaks appear at 0.68 and 0.24, demonstrating uncertainty that would not have been captured with conventional amplitude and frequency extraction methods.

Results are conditional on the considered number of modes present in the data. An analysis is performed including the number of modal arrivals as an unknown in addition to modal amplitudes, frequencies, and noise variance. The presented results correspond to the number of arrivals for which we calculate the highest posterior probability.

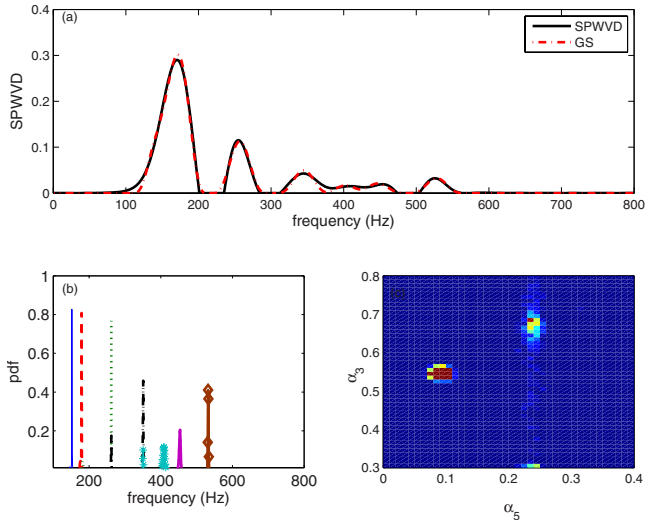


Fig. 3. (Color online) (a) A "slice" from the SPWVD representation of an acoustic signal from the Gulf of Mexico experiment (solid line); the best fit calculated via the Gibbs sampler is superimposed (dot-dashed). (b) The marginal posterior probability distributions for $f_1, \ldots, f_7$. (c) The joint marginal distribution for α_3 and α_5.

5. Conclusions

A new approach is proposed for decomposition of a multimodal signal into components revealing time-frequency and amplitude information for distinct modes. The approach is optimized with a Gibbs sampler that estimates posterior distributions of center frequencies and corresponding amplitudes arriving at a specific time t, and allows quantification of uncertainty in the determination of modal characteristics which can be subsequently employed in geoacoustic inversion interpretation. Our method was applied successfully to synthetic and real data collected during the Gulf of Mexico experiment.

Acknowledgments

This work was supported by the Office of Naval Research through Grant No. N000140510262. The author is grateful to Dr. Jack Ianniello for making the Gulf of Mexico data available.

References and links

[1]G. Okopal, P. J. Loughlin, and L. Cohen, "Dispersion-invariant features for classification," J. Acoust. Soc. Am. **123**, 832–841 (2008).
[2]S. D. Rajan, J. F. Lynch, and G. V. Frisk, "Perturbative inversion methods for obtaining bottom geoacoustic parameters in shallow water," J. Acoust. Soc. Am. **82**, 998–1017 (1987).
[3]C.-S. Chen, J. Miller, F. Boudreaux-Bartels, G. Potty, and C. Lazauski, "Time-frequency representations for wideband acoustic signals in shallow water," in *Proceedings of Oceans 2003* (2003), pp. 2903–2907.
[4]M. Taroudakis and G. Tzagkarakis, "On the use of the reassigned wavelet transform for mode identification," J. Comput. Acoust. **12**, 175–196 (2004).
[5]G. Potty, J. Miller, P. Dahl, and C. Lazauski, "Geoacoustic inversion results from the ASIAEX East China Sea Experiment," IEEE J. Ocean. Eng. **29**, 1000–1010 (2004).
[6]T. C. Yang, "Dispersion and ranging of transient signals in the Arctic Ocean," J. Acoust. Soc. Am. **76**, 262–273 (1984).
[7]I. Zorych and Z.-H. Michalopoulou, "Particle filtering for dispersion curve tracking in ocean acoustics," J. Acoust. Soc. Am. **124**, EL45–EL50 (2008).
[8]J. Candy and D. Chambers, "Internal wave signal processing: A model-based approach," IEEE J. Ocean. Eng. **21**, 37–52 (1996).
[9]Z.-H. Michalopoulou and M. Picarelli, "Gibbs sampling for time-delay and amplitude estimation in underwater acoustics," J. Acoust. Soc. Am. **117**, 799–808 (2005).
[10]L. Cohen, *Time Frequency Analysis: Theory and Applications* (Prentice-Hall, Englewood Cliffs, NJ, 1994).
[11]R. P. Porter, "Transmission and reception of transient signals in a SOFAR channel," J. Acoust. Soc. Am. **54**, 1081–1091 (1973).
[12]S. Dosso, "Quantifying uncertainty in matched field inversion. I. A fast Gibbs sampler approach," J. Acoust. Soc. Am. **111**, 129–142 (2002).
[13]D. J. Battle, P. Gerstoft, W. S. Hodgkiss, W. A. Kuperman, and P. L. Nielsen, "Bayesian model selection applied to self-noise geoacoustic inversion," J. Acoust. Soc. Am. **116**, 2043–2056 (2004).
[14]G. Box and G. Tiao, *Bayesian Inference in Statistical Analysis* (Wiley, Reading, MA, 1973).
[15]S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-6**, 721–741 (1984).
[16]A. E. Gelfand and A. F. Smith, "Sampling-based approaches to calculating marginal densities," J. Am. Stat. Assoc. **85**, 398–409 (1990).
[17]Z. H. Michalopoulou and M. B. Porter, "Focalization in the Gulf of Mexico," in ICASSP-96, Atlanta, GA (1996), Vol. **6**, pp. 3086–3089.
[18]M. B. Porter, "The Kraken normal mode program," Naval Research Laboratory Mem. Report No. 6920, NRL, Washington, DC, 1991.

Relation of sound absorption and shallow water modal attenuation to plane wave attenuation

Allan D. Pierce
Department of Mechanical Engineering, Boston University, Boston, Massachusetts 02215
adp@bu.edu

Abstract: Prediction of attenuation of acoustic fields in weakly absorbing media often uses the substitution of $(\omega/c)\rightarrow(\omega/c)+i\alpha_{\mathrm{pw}}$ into the idealized equations for constant frequency, with α_{pw} representing the local plane wave attenuation coefficient. This assumption is flawed whenever the local absorption of sound is proportional to the square of the gradient of the acoustic pressure, as is the case when the absorption is caused by fluid velocity relaxation. A realistic analysis yields an improved weighting function over depth for determination of guided mode attenuation coefficients.

PACS numbers: 43.30.Es, 43.30.Ma, 43.20.Hq, 43.20.Bi [WS]
Date Received: August 11, 2009 **Date Accepted:** September 18, 2009

1. Introduction

For a homogeneous medium, the linear equations adopted by researchers as governing acoustic propagation invariably admit an approximate solution for plane wave propagation of the form

$$p=\mathrm{Re}\{e^{-i\omega t}\hat{p}(x)\},\quad \hat{p}=Pe^{i(\omega/c)x}e^{-\alpha_{\mathrm{pw}}x}. \tag{1}$$

Here p is the acoustic pressure, $\hat{p}$ is the complex pressure amplitude, P is a constant, ω is the angular frequency, and c is the frequency-independent speed of sound. The approximate validity of this result requires that the frequency-dependent plane wave attenuation coefficient α_{pw} be much less than ω/c. For an inhomogeneous medium, one can define a local parameter $\alpha_{\mathrm{pw}}(\boldsymbol{r})$, depending on spatial position $\boldsymbol{r}$, which would correspond to plane wave propagation in a hypothetical homogenous medium whose material properties, such as sound speed and density, are everywhere the same as those at the specific point $\boldsymbol{r}$. The present paper is concerned with how the absorption of sound in an inhomogeneous medium and the attenuation of guided waves are related to this spatially dependent field $\alpha_{\mathrm{pw}}(\boldsymbol{r})$.

A common assumption, dating at least as far back as to a 1955 paper by Kornhauser and Raney,[1] is that sound absorption can be satisfactorily incorporated into the Helmholtz equation (with time dependence given by the factor $e^{-i\omega t}$) by the replacement $\omega/c\rightarrow(\omega/c)+i\alpha_{\mathrm{pw}}$, so that, with the assumption $\alpha_{\mathrm{pw}}\ll\omega/c$, the Helmholtz equation for propagation in a medium with (possibly) spatially dependent sound speed and density[2] is transformed to an approximate form

$$\rho\,\boldsymbol{\nabla}\cdot\left(\frac{1}{\rho}\boldsymbol{\nabla}\hat{p}\right)+\frac{\omega^2}{c^2}\hat{p}=-2i\alpha_{\mathrm{pw}}\frac{\omega}{c}\hat{p}. \tag{2}$$

The tacit assertion is that this substitution is sufficiently valid, regardless of whatever physical mechanisms account for the attenuation.

For a homogeneous medium, Eq. (2) is adequate for small attenuation for any superposition of propagating plane waves, each having a different direction of propagation. It is questionable that, even for a homogeneous medium, it is applicable for disturbances that cannot be represented as propagating plane wave superpositions, and it is more questionable when the medium is inhomogeneous.

2. Example of an explicit acoustic model involving attenuation

Because it is infeasible to here explore even a representative selection of the many acoustic models that have been introduced into literature to account for sound absorption in various categories of media, attention is focused on one specific (and relatively simple) model for which Eq. (2) is questionable. The model was introduced[3] for low-frequency propagation in sandy/silty marine sediments, but it is illustrative of models in which a fluid velocity relaxation process accounts for energy absorption. The generic equations for this model are here rewritten with several simplifying changes in notation as

$$\rho\frac{\partial \boldsymbol{v}}{\partial t} = -\boldsymbol{\nabla} p, \quad \frac{1}{\rho c^2}\frac{\partial p}{\partial t} + \boldsymbol{\nabla}\cdot\boldsymbol{v} = -\boldsymbol{\nabla}\cdot\boldsymbol{u}, \tag{3}$$

$$\boldsymbol{u} + \tau\frac{\partial \boldsymbol{u}}{\partial t} = \tau_o\frac{\partial \boldsymbol{v}}{\partial t}. \tag{4}$$

The characteristic times τ_o and τ are positive material properties, which may depend on position and are independent of time; τ is substantially larger than τ_o. (That the above is the appropriate form for an inhomogeneous medium may be verified by tracing through the original derivation of these equations.) If the quantity τ_o is set to zero, the solution for the auxiliary velocity quantity $\boldsymbol{u}$ becomes zero, and one recovers the usual acoustic equations for a medium without any dissipation. The physical interpretations of the quantities $\boldsymbol{u}$, τ, and τ_o are not relevant to the present paper, so a discussion is omitted in the interests of brevity.

With regard to sound absorption, the above set of equations has the exact energy corollary

$$\frac{\partial \mathcal{E}}{\partial t} + \boldsymbol{\nabla}\cdot\boldsymbol{I} = -\mathcal{D}, \tag{5}$$

with energy density, intensity, and energy dissipation rate identified as

$$\mathcal{E} = \frac{1}{2}\rho\left(v^2 + \frac{\tau}{\tau_o}u^2\right) + \frac{1}{2\rho c^2}p^2, \tag{6}$$

$$\boldsymbol{I} = p(\boldsymbol{v}+\boldsymbol{u}), \quad \mathcal{D} = \frac{1}{\tau_o}\rho u^2. \tag{7}$$

The quantity $\mathcal{D}$ is the energy dissipated per unit time and per unit volume, and it is always positive. (A natural question is what happens in these expressions when $\tau_o \to 0$. The apparent singularity is illusory, as the quantity u^2/τ_o must go to zero in this limit.)

For disturbances of constant frequency, the complex amplitude of the internal variable $\boldsymbol{u}$ (proportional to slippage velocity) is related to that of the gradient of pressure by

$$\hat{\boldsymbol{u}} = -\left(\frac{\tau_o}{1-i\omega\tau}\right)\frac{1}{\rho}\boldsymbol{\nabla}\hat{p}, \tag{8}$$

so the time average of the energy dissipation per unit time and volume is

$$(\mathcal{D})_{\mathrm{av}} = \frac{1}{2}\left(\frac{\tau_o}{[1+(\omega\tau)^2]}\right)\frac{1}{\rho}\boldsymbol{\nabla}\hat{p}\cdot\boldsymbol{\nabla}\hat{p}^*. \tag{9}$$

The plane wave attenuation constant derived for this model, given all the quantity restraints stated above, is

$$\alpha_{\rm pw} = \frac{\omega^2 \tau_o}{2c[1+\omega^2(\tau+\tau_o)^2]}, \tag{10}$$

and it is also consistent to neglect τ_o relative to τ in the denominator. With the latter simplification, one has

$$(\mathcal{D})_{\rm av} = \left(\frac{c\alpha_{\rm pw}}{\omega^2}\right)\frac{1}{\rho}\nabla\hat{p}\cdot\nabla\hat{p}^*. \tag{11}$$

In contrast, if one traces through a comparable derivation for the heuristic model represented by Eq. (2), one arrives at the conclusion that

$$(\mathcal{D})_{\rm av} = \frac{\alpha_{\rm pw}}{\rho c}|\hat{p}|^2. \tag{12}$$

Both equations give the same result for a propagating plane wave, but there are substantially different predictions for standing waves at nodes and antinodes.

The existence of a discrepancy between these two expressions for the time averaged energy dissipation per unit time and volume is supported by the observation that the inhomogeneous wave equation resulting from the physical model described above is

$$\rho\nabla\cdot\left(\frac{1}{\rho}\nabla\hat{p}\right)+\frac{\omega^2}{c^2}\hat{p} = i\rho\nabla\cdot\left(\left[\frac{\omega\tau_o}{1-i\omega\tau}\right]\frac{1}{\rho}\nabla\hat{p}\right). \tag{13}$$

If one keeps only the real part of the quantity that appears in brackets (as is consistent with an approximate determination of wave attenuation), this becomes

$$\rho\nabla\cdot\left(\frac{1}{\rho}\nabla\hat{p}\right)+\frac{\omega^2}{c^2}\hat{p} = \frac{2i\rho}{\omega}\nabla\cdot\left(\alpha_{\rm pw}\frac{c}{\rho}\nabla\hat{p}\right). \tag{14}$$

In general appearance, the right side of this is considerably different from that of Eq. (2).

3. Iteration solution

A possible argument as to why Eq. (2) should be a good approximation to Eq. (14) proceeds as follows. If $\alpha_{\rm pw}$ and c are sufficiently slowly varying with position, a suitable approximation is

$$\rho\nabla\cdot\left(\frac{1}{\rho}\nabla\hat{p}\right)+\frac{\omega^2}{c^2}\hat{p} = \frac{2i\alpha_{\rm pw}c}{\omega}\left\{\rho\nabla\cdot\left(\frac{1}{\rho}\nabla\hat{p}\right)\right\}. \tag{15}$$

The right side of this equation is a perturbation and presumed small, so setting the left side to zero is a good zeroth order approximation. Consequently, one can argue that any modification of the right side, which makes use of the zeroth order equation, is a good first order approximation. Within the right side, one sets

$$\left\{\rho\nabla\cdot\left(\frac{1}{\rho}\nabla\hat{p}\right)\right\}\rightarrow-\frac{\omega^2}{c^2}\hat{p}, \tag{16}$$

and thereby heuristic equation (2) emerges.

A principal question regarding this procedure is whether moving the quantity $\alpha_{\rm pw}c$ outside of the divergence operator in Eq. (14) has a substantial effect on the quantitative results. For example, what would be the effect of doing this if the plane wave attenuation coefficient should change abruptly, as would be the case in shallow water at a water-sediment interface?

4. Guided modes

The question posed above can be addressed in a simple fashion by considering the propagation of a single guided mode in a medium where the sound speed c and density ρ vary with only the z-coordinate. In the absence of attenuation and cylindrical spreading, a possible propagation of the nth guided mode is described by

$$\hat{p} = Ke^{ik_n x}\Psi_n(z). \tag{17}$$

Here k_n is an eigenvalue, and $\Psi_n(z)$ is an eigenfunction that satisfies the ordinary differential equation

$$\rho\frac{d}{dz}\left(\frac{1}{\rho}\frac{d\Psi_n}{dz}\right) + \left(\frac{\omega^2}{c^2} - k_n^2\right)\Psi_n = 0. \tag{18}$$

The boundary conditions depend on the problem being considered. Here a specific example is required, so the generic problem of shallow water propagation is considered, with the z-axis taken as extending vertically downward from a free-surface at $z=0$ through a water-bottom interface at $z=H$, with the bottom taken as unbounded. The bottom, following the precedent of Pekeris,[4] is taken as also being a fluid. Thus, $\Psi_n=0$ at $z=0$, and $\Psi_n\to 0$ as $z\to\infty$. With these boundary conditions, the eigenvalue problem has the integral corollary

$$\int_0^\infty \frac{1}{\rho}\left(k_n^2\Psi_n^2 + \left[\frac{d\Psi_n}{dz}\right]^2\right)dz = \int_0^\infty \frac{\omega^2}{\rho c^2}\Psi_n^2 dz, \tag{19}$$

and the eigenfunctions satisfy the orthogonality relation

$$\int_0^\infty \frac{1}{\rho}\Psi_n\Psi_m dz = 0 \quad \text{if} \quad n \neq m. \tag{20}$$

For approximate determination of the modal attenuation, an appropriate simple initial assumption is that, given a full set of governing equations that accounts for the physical mechanisms of attenuation, one can rigorously derive the reduced wave equation

$$\rho\,\nabla\cdot\left(\frac{1}{\rho}\nabla\hat{p}\right) + \frac{\omega^2}{c^2}\hat{p} = \{\text{OPER}\}\hat{p}, \tag{21}$$

where the operator $\{\text{OPER}\}$ depends on angular frequency ω and also may involve differential operators, such as the gradient and the divergence. This right side is presumed to be, in some sense, small and is here treated as a perturbation. Each of Eqs. (2) and (14) is a special case of this equation.

To develop the perturbation solution, one expands the complex pressure amplitude $\hat{p}$ in terms of the eigenfunctions, so that

$$\hat{p} = \sum_n a_n\Psi_n(z), \tag{22}$$

where the coefficients a_n depend on the horizontal coordinates. [If only the propagating modes are included, the Ψ_n are not a complete set. A complete set can be fabricated if negative eigenvalues (evanescent modes) are included and if the bottom is assumed to have a rigid or pressure-release termination at some arbitrarily great depth. Whether or not this is done has no effect on the approximate results that follow.] Insertion of this expansion into Eq. (21) and subsequent use of Eq. (18) yields

$$\sum_m (\nabla_H^2 a_m + k_n^2 a_m)\Psi_n = \{\text{OPER}\}\left(\sum_m a_m \Psi_m\right), \tag{23}$$

with ∇_H^2 representing the Laplacian in terms of the horizontal coordinates. Multiplication by $(1/\rho)\Psi_n$ for a specific mode number n, and integration over z, then making use of the orthogonality relation, yields

$$(\nabla_H^2 a_n + k_n^2 a_n)\int_0^\infty \frac{1}{\rho}\Psi_n^2 dz = \sum_m \int_0^\infty \frac{1}{\rho}\Psi_n \{\text{OPER}\}(a_m \Psi_m) dz. \tag{24}$$

The above represents a set of coupled partial differential equations for the coefficients a_n. The zeroth order approximation results with the right side ignored. The lowest order approximation that takes damping into account results when the coupling terms are neglected, but the diagonal terms are retained, so that Eq. (24) reduces to

$$(\nabla_H^2 a_n + k_n^2 a_n)\int_0^\infty \frac{1}{\rho}\Psi_n^2 dz = \int_0^\infty \frac{1}{\rho}\Psi_n \{\text{OPER}\}(a_n \Psi_n) dz. \tag{25}$$

The neglect of the coupling terms is analogous to the "light damping approximation" frequently used[5] in the analysis of the vibrations of multi-degree-of-freedom mechanical systems. Its validity requires that no two modes have closely spaced eigenvalues, but if the magnitudes of the differences $k_n^2 - k_m^2$ are finite and the perturbation sufficiently small, it is expected to be an excellent approximation.

For the special case when the operator {OPER} does not involve any spatial derivatives, as in Eq. (2), one can set

$$\text{Im}\{\text{OPER}\} = -2\alpha_{\text{pw}}\frac{\omega}{c}, \tag{26}$$

as appears on the right side of Eq. (2). Thus, the quantity a_n satisfies the differential equation

$$\nabla_H^2 a_n + k_n^2 a_n = \epsilon a_n - 2ik_n \alpha_{\text{mode},n} a_n. \tag{27}$$

Here $\epsilon = -2k_n \Delta k_n$ is a "small" real number that is of no significant interest. The modal attenuation coefficient $\alpha_{\text{mode},n}$ is identified for this case as

$$\alpha_{\text{mode},n} = \frac{\omega}{k_n}\int_0^\infty W_1(z)\frac{\alpha_{\text{pw}}}{c}dz = \frac{\omega}{k_n}\left\langle \frac{\alpha_{\text{pw}}}{c}\right\rangle_1, \tag{28}$$

where the weighting function $W_1(z)$ is

$$W_1(z) = \frac{(1/\rho)\Psi_n^2}{\int_0^\infty (1/\rho)\Psi_n^2 dz}. \tag{29}$$

The subscript "1" that appears here is to distinguish this weighting function from a second one that emerges further below.

Allowing for minor differences in notation, the expression above for the modal attenuation coefficient can be found in various papers in the recent literature. Examples include its being given in implicit forms in papers by Kornhauser and Raney[1] and Ingenito,[6] and in more explicit forms in papers by Rajan *et al.*,[7] and Evans and Carey.[8]

If the governing equations are taken as those of Eq. (3), and if the reduced wave equation is taken as Eq. (14), then

$$\{\mathrm{OPER}\}(a_n\Psi_n) = \frac{2i}{\omega}\left[\alpha_{\mathrm{pw}}c\Psi_n\nabla_H^2 a_n + \rho a_n \frac{d}{dz}\left(\frac{\alpha_{\mathrm{pw}}c}{\rho}\frac{d\Psi_n}{dz}\right)\right]. \tag{30}$$

It is also a sufficient approximation, using the iteration procedure discussed above, to replace $\nabla_H^2 a_n$ with $-k_n^2 a_n$ in this expression on the right side. A brief derivation similar to that leading to Eq. (28), and which makes use of Eq. (19) then yields

$$\alpha_n = \frac{\omega}{k_n c^2}\int_0^\infty (\alpha_{\mathrm{pw}}c)W_2(z)dz = \frac{\omega}{k_n c^2}\langle\alpha_{\mathrm{pw}}c\rangle_2, \tag{31}$$

where the second weighting function is given by

$$W_2(z) = \frac{(1/\rho)[k_n^2\Psi_n^2 + (d\Psi_n/dz)^2]}{\int_0^\infty (1/\rho)[k_n^2\Psi_n^2 + (d\Psi_n/dz)^2]dz}. \tag{32}$$

The two weighting functions differ markedly at any depth z_o where the eigenfunction has a node, so that $\Psi_n(z_o)=0$. At such a depth $W_1(z_o)=0$ and $W_2(z_o)\neq 0$.

5. Partial reconciliation for special case of homogeneous bottom

The two predictions for the modal attenuation are directly proportional to each other for the special case when the depth region $z>H$ is homogeneous in density and sound speed, and when the plane wave attenuation $\alpha_{\mathrm{pw}}(z)$ is nonzero only in this region. In such a case, for a propagating mode, one has, for $z>H$,

$$\Psi_n(z) = De^{-\gamma_n z}, \quad \gamma_n = \left(k_n^2 - \frac{\omega^2}{c_b^2}\right)^{1/2}, \tag{33}$$

where D is a constant, and c_b (b for bottom) is the sound speed for $z>H$. In such circumstances,

$$k_n^2\Psi_n^2 + \left(\frac{d\Psi_n}{dz}\right)^2 = \left(2k_n^2 - \frac{\omega^2}{c_b^2}\right)\Psi_n^2 \quad \text{for} \quad z>H. \tag{34}$$

Consequently, with the use of Eq. (19), one concludes that

$$\frac{W_2(z)}{W_1(z)} = \frac{2k_n^2 c_b^2}{\omega^2} - 1 \quad \text{for} \quad z>H. \tag{35}$$

Because $k_n^2 c_b^2 > \omega^2$, the ratio is always greater than unity, and the weighting according to the heuristic substitution will yield a smaller modal attenuation coefficient than that according to the fluid velocity relaxation model.

An inference from this special case is that the bulk of the geoacoustic inversions reported in literature that make use of the substitution $(\omega/c)\rightarrow(\omega/c)+i\alpha_{\mathrm{pw}}$ tend to overestimate the values of α_{pw} in the sediment. The circumstances where the estimates would be most suspect are those where the eigenfunction has an additional node below the interface $z=H$ between the water column and the sediment. An estimate of the error can be made by taking the weighting function to be $W_2(z)$ instead of $W_1(z)$. Because the fluid relaxation model has a stronger physical basis, the weighting function W_2 is recommended.

Acknowledgments

The author thanks the Ocean Acoustics Program of the U.S. Office of Naval Research for partial support of the research reported here. He also thanks William M. Carey, Richard B. Evans, and James F. Lynch for helpful discussions. He is especially thankful to Stephen V. Kaczkowski for pointing out an algebraic error in an earlier version of this manuscript.

References and links

[1]E. T. Kornhauser and W. P. Raney, "Attenuation in shallow-water propagation due to an absorbing bottom," J. Acoust. Soc. Am. **27**, 689–692 (1955).
[2]P. G. Bergmann, "The wave equation in a medium with a variable index of refraction," J. Acoust. Soc. Am. **17**, 329–333 (1946).
[3]A. D. Pierce and W. M. Carey, "Low-frequency attenuation of acoustic waves in sandy/silty marine sediments," J. Acoust. Soc. Am. **124**, EL308–EL312 (2008).
[4]C. L. Pekeris, "Theory of propagation of explosive sound in shallow water," Mem.-Geol. Soc. Am. **27**, 1–117 (1948).
[5]J. H. Ginsberg, *Mechanical and Structural Vibrations* (Wiley, New York, 2001), pp. 271–272.
[6]F. Ingenito, "Measurement of mode attenuation coefficients in shallow water," J. Acoust. Soc. Am. **53**, 858–863 (1973).
[7]S. D. Rajan, J. F. Lynch, and G. V. Frisk, "Perturbative inversion methods for obtaining bottom geoacoustic parameters in shallow water," J. Acoust. Soc. Am. **82**, 998–1017 (1987).
[8]R. B. Evans and W. M. Carey, "Frequency dependence of sediment attenuation in two low-frequency shallow-water acoustic experimental data sets," IEEE J. Ocean. Eng. **23**, 439–447 (1998).

 [DOI: 10.1121/1.3213452] Published Online 19 October 2009

Analysis of pausing behavior in spontaneous speech using real-time magnetic resonance imaging of articulation

Vikram Ramanarayanan and Erik Bresch
Department of Electrical Engineering, Speech Analysis and Interpretation Laboratory, University of Southern California, Los Angeles, California 90089
vramanar@usc.edu, bresch@usc.edu

Dani Byrd and Louis Goldstein
Department of Linguistics, University of Southern California, Los Angeles, California 90089
dbyrd@usc.edu, louisgol@usc.edu

Shrikanth S. Narayanan
Department of Electrical Engineering and Department of Linguistics, University of Southern California, Los Angeles, California 90089
shri@sipi.usc.edu

Abstract: It is hypothesized that pauses at major syntactic boundaries (i.e., grammatical pauses), but not ungrammatical (e.g., word search) pauses, are planned by a high-level cognitive mechanism that also controls the rate of articulation around these junctures. Real-time magnetic resonance imaging is used to analyze articulation at and around grammatical and ungrammatical pauses in spontaneous speech. Measures quantifying the speed of articulators were developed and applied during these pauses as well as during their immediate neighborhoods. Grammatical pauses were found to have an appreciable drop in speed at the pause itself as compared to ungrammatical pauses, which is consistent with our hypothesis that grammatical pauses are indeed choreographed by a central cognitive planner.

PACS numbers: 43.70.Fq, 43.72.Ar [DO]
Date Received: February 17, 2009 **Date Accepted:** July 30, 2009

1. Pauses during spontaneous speech

Pausing in natural speech can be considered from a listener perspective—how do pauses aid or impair speech understanding—or from a speaker perspective—how do pauses reflect the speech planning process, either operating well or encountering difficulties. In this paper, we use real-time magnetic resonance imaging (MRI) to provide a noninvasive view of the entire length of the moving vocal tract and to examine pauses from a speech production perspective. Pauses can be broadly categorized into planned or grammatical pauses and unplanned or ungrammatical pauses. (For our purposes, we make no distinction between planned and grammatical, and likewise unplanned and ungrammatical pauses.) Grammatical pauses generally occur at the boundary of a clause, presumably due to the need to parse and plan the sentence. Ungrammatical pauses can indicate a breakdown in composing the speech stream and occur at inappropriate locations as the planning, production, and/or lexical access process is disrupted (see O'Shaughnessy, 1992, Rochester, 1973).

The framework of articulatory phonology (Browman and Goldstein, 1992, 1995) in conjunction with the prosodic-gesture model (Byrd and Saltzman, 2003) of phrase boundaries offers one approach for considering the nature of grammatical and ungrammatical pauses in articulation. In this framework, the act of speaking is decomposable into a.u. of vocal tract action—gestures—that can be defined as an equivalence class of goal-directed movements, such as those by a set of articulators in the vocal tract (see the task dynamics model, Saltzman

and Munhall, 1989). Byrd and Saltzman (2003) viewed phrase junctures as phonologically planned intervals of controlled local slowing of speech timing around a phrase edge, with the articulatory slowly increases as the boundary approaches and the speech stream resumes speed as the boundary recedes (i.e., immediately postboundary). This "clock" slowing, at its extreme, can be understood to result in a pause, as the clock controlling articulation slows to a near-stop and then speeds up again as the postpause interval is initiated. In contrast, ungrammatical pauses (which may be filled or unfilled depending on the state of voicing) abruptly interrupt the execution of the planned speech stream interfering with the vocal tract articulators reaching their targets. Under this approach, a grammatical pause, then, is viewed as a planned event under cognitive control with explicit consequences for the spatiotemporal behavior of the articulators over an interval; consequences that are distinct from ungrammatical pauses that abruptly perturb articulation. In this paper, we will examine direct articulatory evidence for this hypothesis.

Although pauses can contribute information about speech planning, few joint acoustic and articulatory studies of pausing behavior have been carried out, one reason being the difficulty of acquiring data on vocal tract movement during running speech. Recent progress in real-time MRI (Narayanan *et al.*, 2004) allows for a more comprehensive investigation of pauses in speech than does study of the acoustic signal alone. Since the technique allows for a complete view of the moving vocal tract, providing synchronized audio in conjunction, it is possible to examine the supraglottal articulators during not only the spoken portion but also the silent portions of the speech stream.

2. Data acquisition and preparation

The data we examined comprise spontaneous speech utterances and the corresponding time-synchronized movies of the moving vocal tract, elicited in response to queries from the experimenter. Seven healthy native speakers of American English were asked to answer simple questions on general topics such as "what music do you listen to…," "tell me more about your favorite cuisine…," etc.) while lying inside a MRI scanner. For each of the stimulus questions, time-synchronized audio responses and MRI videos of speech articulation were recorded for 30 s. Further details regarding the recording/imaging setup can be found in Narayanan *et al.*, 2004; Bresch *et al.*, 2006. Midsagittal real-time MR images of the vocal tract were acquired with a MR pulse repetition time of TR=6.5 ms on a GE Signa 1.5 T scanner with a 13 interleaf spiral gradient echo pulse sequence. The slice thickness was approximately 3 mm. A sliding window reconstruction at a rate of 22.44 frames/s was employed. The field of view was adjusted depending on the subject's head size, so that images covered an area of 18.4×18.4 cm^2 at a resolution of 68×68 pixels.

For the manual annotation of the audio waveform for this experiment, a grammatical pause was defined to be a silent or filled pause that occurred between overt syntactic constituents (including sentence end). Examples include pauses at (1) clause boundaries such as relative clause boundaries, (2) subject-verb or verb-object boundaries, and (3) prepositional phrases offset from another constituent. Any pause other than the above, i.e., generally those occurring within a clause, was marked as an ungrammatical pause. Such pauses are atypical in this natural speech and do not mark the juncture between obvious syntactic or semantic word groups in the sentence; they do not appear to encode linguistic information. For each speaker's utterances, grammatical and ungrammatical pauses were manually annotated by the first author according to this definition and verified by a linguist for accuracy.

3. Analyses

In order to examine the articulatory characteristics at and around pauses, the extraction of a "gradient energy" measure that captures the speed of articulatory motion (of all articulators) from image sequences is employed. In order to study the time evolution of vocal tract shaping, for each set of image sequences, the air-tissue boundary of the articulatory structures needs to

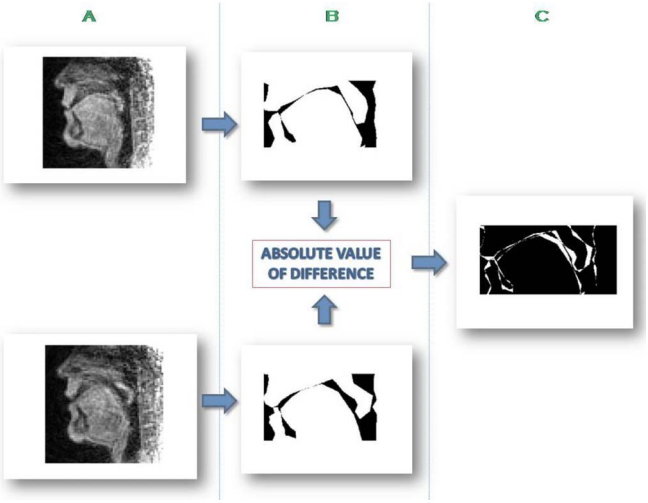


Fig. 1. (Color online) An illustration of the gradient energy calculation process: first, contour outlines are obtained from the MRI images in panel A and are then converted to binary masks (panel B); these are then used to compute the "gradient" images (panel C), the energy of which is then calculated by a simple addition operation (of all white pixels).

be clearly delineated. This contour tracing process is time consuming and tedious when carried out by a human, so an algorithm using Fourier region segmentation to automatically carry out the task was used (see Bresch and Narayanan, 2009).

In order to observe articulatory effects of pausing behavior more comprehensively, it is important to study articulator dynamics not only in the pause frames but also in the interval preceding and following the pause, particularly since models such as the prosodic-gesture model (Byrd & Saltzman, 2003) predict spatiotemporal effects during neighboring intervals. Since most appreciable effects, including construction of a rough plan for the utterance (Kochanski *et al.*, 2003), occur in a time window of 500 ms before and after the pause, neighborhoods of that order were analyzed for global range of movements of articulators. Since in our experimental setup, the frame rate is about 22.44 frames/s; this approximately translates to neighborhoods consisting of about 12 frames. Thus, for analysis, although the length (in number of frames) of each pause was variable, the analysis neighborhoods before and after the pause were of fixed lengths.

Once the contour outlines have been extracted from the MR images, they are used to create binary mask images, with all pixels enclosed by these contour outlines assigned a normalized value of 1, and the rest, 0, such that the midsagittal section of the vocal tract appears white on a black background (see Fig. 1). A gradient energy measure was calculated for every pair of *contiguous mask images* in a pause/neighborhood frame sequence, by subtracting them, taking the absolute value of the difference, and computing the "pixel energy" of the result (by finding the number of pixels of value "1"). The overall gradient energy value for a pause/neighborhood is then computed by averaging over all gradient energies obtained during the pause/neighborhood period. This is done to obtain an entropy measure that can capture variability in articulator movement and thus give an estimate of the speed of articulatory motion during such periods, which this measure does well on a global level.[1] In a similar manner, one can compute delta gradient measures that will capture the acceleration of the articulators during pauses/neighborhoods.

A two-factor parametric analysis of variance (ANOVA) was conducted on the dependent variable of gradient energy with data pooled across speakers and with the factors: site (levels: prepause, pause, and postpause) and grammaticality (levels: grammatical and ungrammatical). It should be noted that since the number of occurrences of grammatical and ungrammatical pauses (especially the latter[2]) were too small for a repeated measures ANOVA, analyses

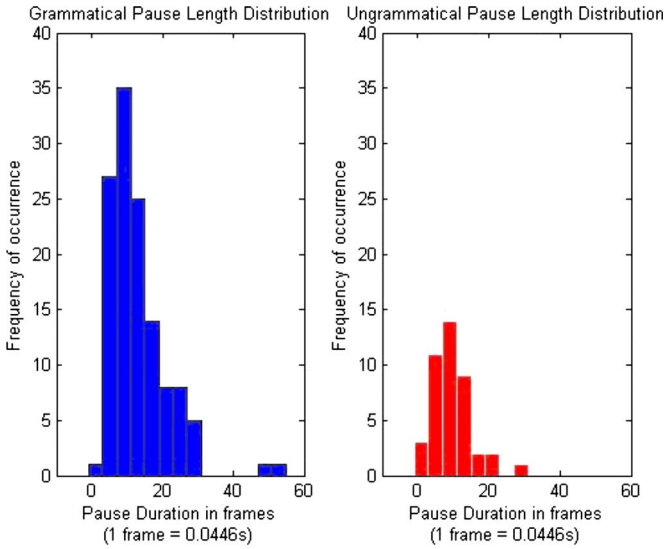


Fig. 2. (Color online) Pause length distributions for grammatical and ungrammatical pauses.

were carried out with data pooled across speakers.[3]

4. Results

Histograms of grammatical and ungrammatical pause durations across all speakers were plotted, and while these pauses cannot be reliably separated based on duration values alone, a statistical t-test indicated that grammatical pauses tended to be significantly longer on average ($p \leq 0.01$)—the mean and standard deviation of pause durations were found to be 13.62 frames and 8.2 frames, respectively, for grammatical, and 9.76 frames and 5.8 frames, respectively, for ungrammatical pauses (1 frame = 0.0446 s) (see Fig. 2).

The time-normalized average gradient frame energy for each pause and for the neighborhoods before and after it, pooled across all 7 speakers, is plotted as a bar graph in Fig. 3. Corresponding time-normalized average local phone rates are also plotted to the right of each of these graphs. The derived measure of mean gradient frame energy captures articulator speeds well and is a good indicator of the local phone rate (assuming that articulator speeds directly inform the local phone rate to a certain extent).

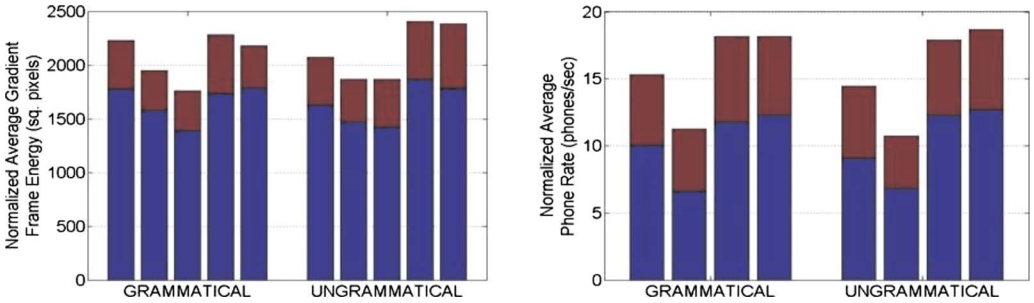


Fig. 3. (Color online) Time-normalized average *gradient frame energies* (in squared pixels) of grammatical and ungrammatical pauses and their neighborhoods pooled across all seven speakers (standard deviation bars are plotted on top of each energy bar in a lighter color). Corresponding average *local phone rates* (phones/s) are also shown to the right of the gradient frame energy panel. Each panel consists of two pause groups on the *x*-axis: (1) grammatical and (2) ungrammatical. Group 1 consists of, in order, bars for two neighborhoods immediately before the grammatical pause (~250 ms), followed by one bar for the pause itself (*not shown for phone rate graph*), followed by two bars for the neighborhoods following the pause (~250 ms); this set of five bars is followed by a parallel sequence of five bars for the ungrammatical pauses (Group 2).

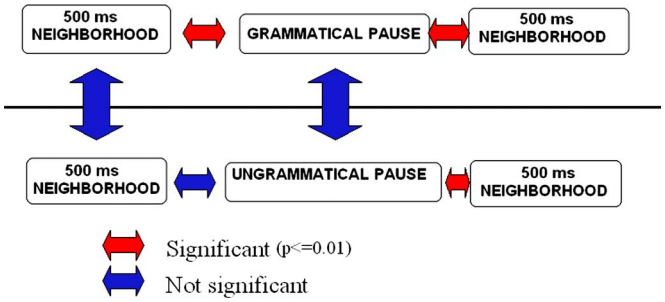


Fig. 4. (Color online) A schematic depicting the levels (grammatical and ungrammatical) and sites (prepause, pause, and postpause) at which the ANOVA statistical analyses were performed.

In order to examine the effects of speech planning on the structure of pauses, we have to examine how the gradient frame energies vary moving into and out of the pause. Figure 4 schematically summarizes the statistical comparisons (double-headed arrows) performed on the data. Significant differences ($p \leq 0.01$) were found between the means of the gradient energies in the (prepausal) neighborhood before *grammatical* pauses and those of the pauses themselves. That is, the gradient energy means of the grammatical pauses themselves were *significantly lower* than these prepausal neighborhood gradient energy means. In contrast, *ungrammatical* pauses displayed *no significant differences* between the means of the prepausal and pausal gradient energies. However, there was a significant increase in the gradient energy for the neighborhood immediately following both grammatical and ungrammatical pauses, which was often slightly higher than the prepausal energy value. There were no significant differences in the means of the grammatical and ungrammatical (i) pause gradient energies or (ii) prepausal neighborhood gradient energies. However, ungrammatical pauses showed a significantly higher variation in the values of gradient energies compared to the grammatical case, especially during and after the pause (see Table 1), which is expected, since such a pause would hypothetically serve to interrupt the flow of speech (irrespective of the speech rate) and hence would have a much higher gradient energy variance compared to grammatical pauses.

For both grammatical and ungrammatical pauses, there was no trend found that distinguished filled from unfilled pauses, as the gradient energies of these cases can be highly context dependent. In some cases, the filled pause gradient energies for some speakers were much higher than their unfilled counterparts, while the opposite effect was found for other speakers. Furthermore, there was no observed trend of gradient frame energy variation *within* a pause, be it grammatical or ungrammatical, filled or unfilled, suggesting that instantaneous values of these gradient energies may be context dependent.

The results obtained suggest that grammatical pauses are part of a more globally choreographed plan of articulatory movement, since at the pause, the speed of the articulators drops (as indicated by the reduction in mean gradient energy), which, finally, toward the end, increases to around the level where it was at the start of the pause. Also, the results suggest that ungrammatical pauses are essentially unplanned, with the articulator speed dropping slightly (but not significantly) early into the pause, following which there is a sudden jump in the gradient en-

Table 1. Standard deviation (in squared pixels) of the gradient energies for grammatical and ungrammatical pauses and their neighborhoods pooled across all speakers (here the two 250 ms neighborhoods before and after the pause are pooled together to get one 500 ms neighborhood before and after).

Grammatical pauses			Ungrammatical pauses		
500 ms before	Pause	500 ms after	500 ms before	Pause	500 ms after
366.54	369.25	423.57	374.95	445.37	538.11

ergy. In addition, there is a large variance associated with the gradient energy values in this case (Table 1). Such pauses in spontaneous speech, which occur without the linguistic structuring of the speaker, are characterized by a sudden increase in articulator speed on a global level when the speaker eventually succeeds in lexical access or planning.

5. Conclusions

In this paper, our principal hypothesis that grammatical pauses are a result of a higher-level cognitive plan of articulatory movement, while ungrammatical ones are not, has been validated via direct observation of articulatory behavior. Measures that help distinguish between the two in the articulatory domain were developed.

It has long been recognized that pauses are relevant to cognitive processing and are related to effect, style, and lexical and grammatical structure (e.g., Lehiste, 1970; Rochester, 1973; Kutik *et al.*, 1983; Zellner, 1994). Direct observation of articulation along the entire vocal tract offers an important new source of data for investigation of speech planning, since it allows a view of how the speech flow is altered in either a cognitively planned way or interrupted by a perturbation when normal planning fails. It can also inform as to how much time it takes to "recover" from the effect of a sudden unplanned pause that perturbs the linguistic structural integrity of the utterance.

Acknowledgments

The authors thank Ed Holsinger and Krishna Nayak. Work described in this paper was supported by NIH Grant Nos. DC007124 and DC03172, the USC Imaging Sciences Center, and the USC Center for High Performance Computing and Communications (HPCC).

References and links

[1]By "global," we mean that the gradient energy measure is an indicator of the net motion of all vocal tract articulators.

[2]Some utterances of some speakers were found to not contain any ungrammatical pauses.

[3]Also, due to the same reason, we cannot directly assume that the data are parametric, although the results obtained using a nonparametric data distribution assumption are similar to those obtained using parametric analysis, and so only the latter results are reported.

Bresch, E. and Narayanan, S. (**2009**). "Region segmentation in the frequency domain applied to upper airway real-time magnetic resonance images," IEEE Trans. Med. Imaging **28**, 323–338.

Bresch, E., Nielsen, J., Nayak, K., and Narayanan, S. (**2006**). "Synchronized and noise-robust audio recordings during realtime MRI scans," J. Acoust. Soc. Am. **120**, 1791–1794.

Browman, C. P. and Goldstein, L. (**1992**). "Articulatory phonology: An overview," Phonetica **49**, 155–180.

Browman, C. P. and Goldstein, L. (**1995**). "Dynamics and articulatory phonology," in *Mind as Motion: Dynamics, Behavior, and Cognition*, edited by R. Port and T. van Gelder (MIT, Cambridge, MA), pp. 175–193.

Byrd, D. and Saltzman, E. (**2003**). "The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening," J. Phonetics **31**, 149–180.

Kochanski, G., Shih, C., and Jing, H. (**2003**). "Quantitative measurement of prosodic strength in Mandarin," Speech Commun. **41**, 625–645.

Kutik, E. J., Cooper, W. E., and Boyce, S. (**1983**). "Declination of fundamental frequency in speakers' production of parenthetical and main clauses," J. Acoust. Soc. Am. **73**, 1731–1738.

Lehiste, I. (**1970**). *Suprasegmentals* (MIT, Cambridge, MA)

Narayanan, S., Nayak, K., Lee, S., Sethy, A., and Byrd, D. (**2004**). "An approach to real-time magnetic resonance imaging for speech production," J. Acoust. Soc. Am. **115**, 1771–1776.

O'Shaughnessy, D. (**1992**). "Recognition of hesitations in spontaneous speech," in Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing, San Francisco, CA, pp. 521–524.

Rochester, S. R. (**1973**). "The significance of pauses in spontaneous speech," J. Psycholinguist. Res. **2**, 51–82.

Saltzman, E. L. and Munhall, K. G. (**1989**). "A dynamical approach to gestural patterning in speech production," Ecological Psychol. **1**, 333–382.

Zellner, B. (**1994**). "Pauses and the temporal structure of speech," in *Fundamentals of Speech Synthesis and Speech Recognition*, edited by E. Keller (Wiley, Chichester), pp. 41–62.

Acoustic coupling between pistons in a rigid baffle

Kassiani Kotsidou
Saginaw Valley State University, University Center, Michigan 48710
kkotsido@svsu.edu

Charles Thompson
UMass Lowell, Lowell Massachusetts 01854
charles_thompson2@uml.edu

Abstract: This paper demonstrates the merit of "matched asymptotic expansions" by applying the method to the analysis of the acoustical coupling between vibrating pistons. The accuracy of the method is verified by comparing the results of this study with existing solutions. The method uses the disparity between the characteristic length scale of the nearly incompressible fluid motion near the piston and that of the far field acoustic pressure. The velocity potential in each region is developed in terms of a singular perturbation expansion. Finally, the combination of the locally valid solutions leads to a global solution.

PACS numbers: 43.38.Hz [RW]
Date Received: August 7, 2008 **Date Accepted:** September 11, 2009

1. Introduction

Acoustic coupling describes the mutual interaction between acoustic sources that vibrate in close proximity. Its study, along with studies of spatial distributed sound sources, originated in the early part of the 20th century. In 1903, Lord Rayleigh[1] examined the problem of enhanced radiation efficiency and appreciated the mutual interaction among distributed sound sources. In 1939, Klapman[2] developed the mutual impedance as it is defined and used today. In 1960, Pritchard[3] examined the mutual radiation impedance by utilizing the complex angle formulation of the inverse Fourier transform with respect to the spatial wavenumber spectrum of the pressure. Pritchard's work concluded with a low frequency approximation, which has found common usage among the underwater acoustics engineers. Scandrett *et al.*[4] examined the accuracy of Pritchard's approximation and concluded that a modified version could be used at low frequency for highly dense transducer arrays.

The approach taken here differs from prior work in that it considers the characteristics of the fluid motion in the development of the field description and the coupling that ensues. Our objective here is to provide a general approach that will allow one to extend the basic result presented. The problem domain will be split into wave and incompressible regions. In each region the velocity potential will be expressed in terms of locally valid asymptotic expansions in space. Finally, the expansions are matched[5–7] to obtain a solution globally valid in space. Section 2 will present the problem statement, will outline the adopted approach, and will summarize our findings. The common problem of two circular pistons in an infinite baffle will allow us to compare our approach with well known results (Sec. 3). Finally, Sec. 4 will provide a synopsis that concludes this paper. From space considerations, detailed information about the method of matched asymptotic expansions (MAEs) will not be provided. The interested reader may use sources like Refs. 8 and 6.

2. Evaluation of the pressure field

Consider two vibrating pistons placed in a rigid baffle. The radius of each piston is given by a_1H_0 and a_2H_0, where H_0 is the characteristic radius of the pistons. A depiction of the problem

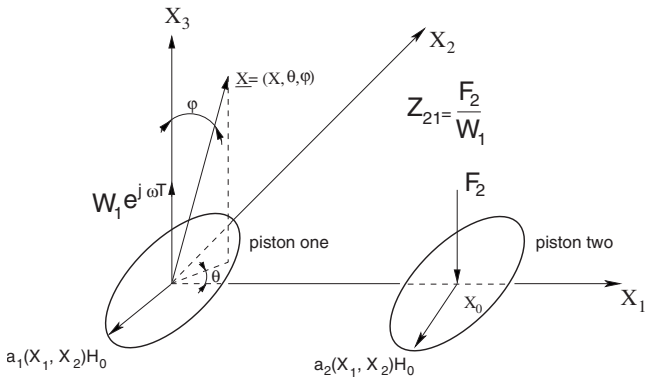


Fig. 1. Schematic of two pistons in a rigid baffle.

geometry is given in Fig. 1. Each piston is located in a rigid baffle which spans the X_1-X_2 plane. The center of the first piston is located at the origin of the coordinate system, whereas the second piston is located at $(X_0, 0, 0)$.

The first piston is driven into time harmonic motion with frequency ω and vibrates with a complex velocity amplitude $W_1 = W_0 w_1$, where W_0 is the typical velocity amplitude. A schematic of the problem geometry is given in Fig. 1. The variable X is the spherical radius, θ is the azimuth angle, and φ is the zenith angle. The motion of the first piston gives rise to a pressure wave, and as a result, to a force F_2 on the second piston. Blocking the motion of the second piston will allow the evaluation of this force. The interaction between the pistons is described via the mechanical mutual impedance Z_{21}, which is defined as the ratio of F_2 over W_1. The task at hand is, therefore, the evaluation of the relationship between the velocity W_1 and the force F_2.

The problem domain may be split into incompressible and wave regions. The region in the vicinity of the pistons is considered to be the incompressible region because the fluid motion is inertially dominated. That is, the inertial forces dominate over the compressive forces. The measure of the balance between these forces is related to the wavelength. The remainder of the domain is termed the wave region, where the wave motion is described by the known wave equation. In the incompressible region, the length scale H_0 is used, which is much less than the wavelength λ. In the wave region, the length scale L_0 is used. This scale is comparable to the wavelength λ. Our analysis focuses on the case where $H_0 \ll L_0$ and the approximation is expressed in term functions of $\epsilon = H_0/L_0$.

In each region, the variables are normalized with respect to the typical parameters of the region. In the wave region, the normalized velocity potential adheres to the known wave equation. Hence, Eq. (1) applies

$$\phi(\underline{x}, \epsilon) = \mathcal{A}(\epsilon)\frac{e^{-jkx}}{x} = \left(\sum_{n=0}^{2} \mathcal{A}_n \epsilon^n\right)\frac{e^{-jkx}}{x} + O(\epsilon^3) = \mathbf{S}_2 \phi(\underline{x}, \epsilon), \tag{1}$$

where $\mathcal{A}(\epsilon)$ has been replaced by its asymptotic series in ϵ and $\mathbf{S}_2$ truncates the series at order ϵ^2. With λ being the wavelength, $k = (2\pi/\lambda)L_0$ and $\underline{x} = x/L_0$ are the wavenumber and the length vector normalized in the wave region. In the incompressible region, the velocity potential is sought in terms of the asymptotic series

$$\mathbf{S}_2 \hat{\phi}(\underline{\hat{x}}, \epsilon) = \sum_{n=0}^{2} \hat{\phi}_n(\underline{\hat{x}})\epsilon^n + O(\epsilon^3). \tag{2}$$

Considering the underlying physics of the problem at hand, the following equations hold:

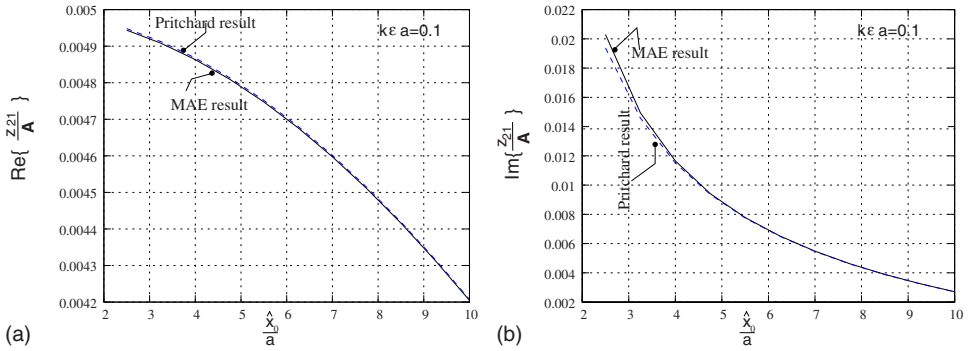


Fig. 2. (Color online) Mutual impedance z_{21} between two pistons versus their distance x_0.

$$\hat{\phi}_0(\underline{\hat{x}}) = -\int\int_{\text{piston}} w_1 \frac{1}{2\pi}\frac{1}{|\underline{\hat{x}}-\underline{\hat{\xi}}|} dS_{\hat{\xi}},$$

$$\hat{\phi}_1 = \text{const},$$

$$\hat{\phi}_2(\underline{\hat{x}}) = \frac{k^2}{4\pi}\int\int_{\text{piston}} w_1 |\underline{\hat{x}}-\underline{\hat{\xi}}| dS_{\hat{\xi}}. \tag{3}$$

In the above equations, $\underline{\hat{x}}=X/H_0$ is the length vector normalized in the incompressible region and $\underline{\hat{\xi}}$ is the integration variable. Applying a matching paradigm, the globally solution for the velocity potential is to be

$$\mathbf{S}_{2,2}\phi = \hat{\phi}_0(\underline{\hat{x}}) + \epsilon^2\hat{\phi}_2(\underline{\hat{x}}) + \frac{\mathcal{A}_1}{\hat{x}}e^{-jk\epsilon\hat{x}} - \frac{\mathcal{A}_1}{\hat{x}}\left(1-\frac{k^2\epsilon^2\hat{x}^2}{2}\right), \tag{4}$$

where $\mathcal{A}_1=-\bar{w}_1\mathbf{A}/2\pi$ and $\mathbf{A}$ is an arbitrary constant.

3. Mutual impedance

Combining all the above results, the mutual impedance between the two pistons is given by

$$z_{21} = -jk\epsilon\frac{\int\int_{\text{2nd piston}} S_{m,n}\phi dS}{w_1}. \tag{5}$$

Figures 2(a) and 2(b) compares the approximation set forth by Pritchard to the result obtained by MAE for $ka\epsilon=0.1$. Pritchard's result given in non-dimensional form is

$$\tilde{z}_{21} = \frac{\int\int_{\text{2nd piston}} p(\underline{\hat{x}})dS}{w_1} = jk\epsilon\frac{\pi a_1^4}{2}\frac{e^{-jk\epsilon\hat{x}_0}}{\hat{x}_0} = j(k\epsilon a)\frac{e^{-jk\epsilon a(\hat{x}_0/a_1)}}{\left(\frac{\hat{x}_0}{a_1}\right)}\pi a_1^2 \tag{6}$$

for $(k\epsilon a_1)^2 \ll 1$, $(\hat{x}_0/a_1) \gg 1$, with $\hat{x}_0$ being the distance between the pistons normalized in the incompressible region. For large distance between the pistons, the two results match. As the distance between the pistons is reduced, both the real and the imaginary parts of the current result depart from that of Pritchard. The divergence is more apparent in the imaginary part of the impedance. Nevertheless, Pritchard clearly stated that his approximation is valid when the

distance between the two pistons is much greater than their dimensions. The advantage in the application of MAEs is that one can remove the restriction on the shape and velocity distribution of the piston.

4. Concluding remarks

This paper has demonstrated that the method of MAEs can be used in the calculation of the mutual impedance. The primary technique that has been employed in the past and is currently employed uses the classical wave theory. Nevertheless, the classical wave theory is confined to the geometry and the velocity distribution of the problem. On the other hand, the technique of matched asymptotic expansions is a simplified approach not confined to circular pistons or uniform velocity distribution. The method of MAE has been demonstrated by an example whose exact solution is known. The MAE outcome was compared with the exact solution for well separated sources. In the case that the wavelength is large compared to the dimensions of the piston, 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 a function of the time of flight between a transducer pair.

References and links

[1]L. Rayleigh, "On the production and distribution of sound," Philos. Mag. **6**, 1061–1065 (1903).
[2]S. J. Klapman, "Interaction impedance of a system of circular pistons," J. Acoust. Soc. Am. **11**, 289–295 (1940).
[3]R. I. Pritchard, "Mutual acoustic impedance between radiators in an infinite rigid plane," J. Acoust. Soc. Am. **32**, 730–737 (1960).
[4]C. L. Scandrett, J. L. Day, and S. R. Baker, "A modal Pritchard approximation for computing array element mutual impedance," J. Acoust. Soc. Am. **109**, 2715–2729 (2001).
[5]J. R. O'Malley, *Introduction to Singular Perturbations*, Applied Mathematics and Mechanics, Vol. **14** (Academic, New York, 1974), Chaps. 1 and 2.
[6]C. Thompson, "Linear inviscid wave propagation in a waveguide having a single boundary discontinuity: Part I: Theory," J. Acoust. Soc. Am. **75**, 346–355 (1984).
[7]B. Shivamoggi, *Perturbation Methods for Differential Equations*, 1st ed. (Birkhauser, Boston, 2002), Chap. 5.
[8]M. Lesser and D. Crighton, "Physical acoustics and the methods of matched asymptotic expansions," in *Physical Acoustics*, edited by W. Mason and R. Thurston (Academic Press, New York, 1975), Vol. **11**.

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

High-rate envelope information in many channels provides resistance to reduction of speech intelligibility produced by multi-channel fast-acting compression (L)

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

(Received 5 May 2009; revised 18 August 2009; accepted 23 August 2009)

The intelligibility of speech in a competing-speech background was measured for signals that were subjected to multi-channel compression and then tone vocoded. The lowpass filter used to extract the envelopes in the vocoder preserved only low-rate envelope cues (E filter) or also preserved pitch-related cues (P filter). Intelligibility worsened with increasing number of compression channels and compression speed, but this effect was markedly reduced when the P filter was used and the number of vocoder channels was 16 as compared to 8. Thus, providing high-rate envelope cues in many channels provides resistance to the deleterious effects of fast compression.
 [DOI: 10.1121/1.3238159]

PACS number(s): 43.66.Ts, 43.71.Gv, 43.66.Mk [RLF] Pages: 2155–2158

I. INTRODUCTION

Fast-acting dynamic range compression is used in hearing aids and cochlear implants to reduce short-term fluctuations in the level of the signal so as to maintain audibility and comfort. The reduction in envelope modulation produced by the compression increases with decreasing envelope rate, decreasing attack and release times, and increasing compression ratio (Braida *et al.*, 1982; Stone and Moore, 1992). When the compression is applied independently in several frequency channels, as is usually the case, it also leads to reduced spectral contrast (Plomp, 1988). Reduction in temporal and/or spectral contrast can lead to reduced speech intelligibility, especially when background sounds are present (Baer and Moore, 1994; Chi *et al.*, 1999). Conversely, the improved audibility of low-level signal components produced by multi-channel fast-acting compression sometimes leads to improved intelligibility (Moore *et al.*, 1992; Yund and Buckles, 1995; Moore *et al.*, 1999), but this does not always occur, especially for listeners with severe or profound hearing loss (HL) (De Gennaro *et al.*, 1986; Boothroyd *et al.*, 1988).

Stone and Moore (2008) investigated how the number of compression channels and their speed influenced performance in a task which required listeners to identify the speech of a target talker in the presence of a background talker. To simulate loss of access to temporal fine structure information as well as the reduced access to spectral detail experienced by people with cochlear hearing loss and people with cochlear implants (Hopkins and Moore, 2007; Moore, 2008), they processed the signals using a noise vocoder with either 12 or 18 channels. A second method of limiting the information conveyed by the vocoded signal was employed: the lowpass filter, used to extract the temporal envelope in each channel of the vocoder, had a cutoff frequency of 50 Hz and a slope of −24 dB/octave. This was intended to remove envelope information related to the fundamental frequency (f0), while preserving low-rate envelope cues related to movement of the articulators (Rosen, 1992). Stone and Moore (2008) showed that as the number of channels of compression and/or their speed increased, intelligibility in the competing-speech task decreased: surprisingly, even relatively slow compression reduced intelligibility.

As noted above, the processing used by Stone and Moore (2008) removed f0-related envelope cues since channel envelopes were lowpass filtered at 50 Hz to ensure negligible response by 100 Hz, the lower end of the range of f0 used by the speakers of the test material. However, envelope cues for rates much higher than this can be used by people with cochlear hearing loss (Bacon and Viemeister, 1985; Moore and Glasberg, 2001) and people with cochlear implants (Shannon, 1992). Therefore, the processing used by Stone and Moore (2008) removed f0-related envelope information that would normally be at least partially available to people with cochlear hearing loss and cochlear implantees.

[a]Author to whom correspondence should be addressed. Electronic mail: mas19@cam.ac.uk

Provision of higher-rate envelope cues in either a tone or noise vocoder usually leads to higher intelligibility, especially when background sounds are present (Whitmal *et al.*, 2007; Stone *et al.*, 2008). Furthermore, when more than five vocoder channels are used, the use of tone carriers as opposed to noise carriers generally produces higher intelligibility, probably because the latter have inherent random modulations (Whitmal *et al.*, 2007; Stone *et al.*, 2008). The present experiment extends the work of Stone and Moore (2008) by assessing the effect of preserving higher-rate envelope modulation in the vocoded signal. Additionally, by using a tone vocoder rather than a noise vocoder, the possible confounding effect of the random modulation introduced by the latter was eliminated.

II. SPEECH MATERIALS AND PROCESSING

Listeners were required to identify target sentences presented in a background of a competing talker. The target speech was taken from the IEEE corpus (IEEE, 1969) and was spoken by a male using British English. The background talker was also a male with a mean f0 that was 0.5 octaves higher than the mean f0 of the target talker (approximately 100 Hz). Further details of both target and background recordings are given in Stone *et al.* (2008).

The processing method was similar to that described by Stone and Moore (2008). The stimuli were first processed using a multi-channel compression system and then processed through a tone vocoder with N_v channels, where $N_v = 8$ or 16. The values of N_v were selected to span the range of the effective number of channels available to cochlear-implant users (Friesen *et al.*, 2001) and to listeners with a moderate-to-severe cochlear hearing loss (Moore, 2007). The number of compression channels N_c was $N_v/8$, $N_v/4$, or $N_v/2$.

The speech was mixed with the background at a fixed target-to-background ratio (TBR) before being processed. For each N_c, three compression speeds were used but the compression speed was the same for all channels. In the vocoder processing, each channel envelope was extracted by half-wave rectification before lowpass filtering. The response of the lowpass filter was −6 dB at either 50 Hz (the envelope or E filter) or 200 Hz (the pitch or P filter). The filter responses were −30 dB by 100 and 400 Hz, respectively. The lowpass-filter cutoff was the same for all channels of the vocoder. Each envelope signal was used to modulate a sinusoidal carrier at the arithmetic center frequency of its vocoder channel. Each modulated carrier was filtered to restrict its spectrum to the frequency range of its respective channel. The filtered modulated carriers were then combined.

TBRs were chosen based on pilot studies so as to produce mean word intelligibility of around 50%–70%. For $N_v = 16$, the TBR was +6 dB when using the E filter and +2 dB when using the P filter. For $N_v = 8$, the TBRs were +10 dB (E filter) and +8 dB (P filter).

The channels for both the compressor and the vocoder processing were equally spaced and had fixed bandwidths on the ERB_N-number scale (Glasberg and Moore, 1990). The channel widths for the 8- and 16-channel vocoders were 3.71 and 1.86 ERB_N, respectively. The 16-channel system had channel edge frequencies of 100, 173, 261, 370, 502, 664, 861, 1102, 1396, 1755, 2194, 2729, 3383, 4181, 5156, 6347, and 7800 Hz. Channel edges for the eight-, four-, two-, and one-channel systems were selected from appropriate subsets of these values. For example, for the four-channel system, the edge frequencies were 100, 502, 1396, 3383, and 7800 Hz. The edge frequencies were used to generate a filterbank that was unique for each channel-number condition, rather than to group the relevant channels from the output of a 16-channel filterbank. Note that the relatively small bandwidths of some of the low-frequency channels for $N_v = 8$ and 16 meant that the highest modulation rate that could be carried by those channels was lower than the limit imposed by the envelope filter (Stone *et al.*, 2008).

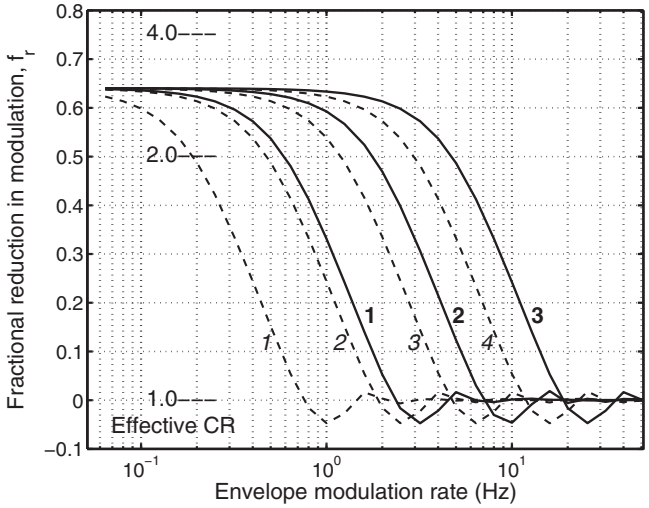


FIG. 1. Plots of fractional reduction in modulation (f_r) as a function of envelope modulation rate (sinusoidal modulation) for the four compression speeds used in experiment 3 of Stone and Moore (2008) (dashed lines and italic numbering) and the three compression speeds used in the present study (solid lines and bold numbering). For ease of conversion from the f_r scale, markers for effective compression ratio ("Effective CR") are also shown.

III. COMPRESSION SPEED AND RATIO

The same "envelope" compressor as used in experiment 3 of Stone and Moore (2008) was used here. The corner frequencies of the two-pole lowpass filter used to extract the signal envelope were 0.56 (speed 1), 1.68 (speed 2), and 4.48 Hz (speed 3). The resulting, near-symmetric, attack and release times were 448, 149, and 56 ms, respectively, corresponding to moderately slow, through syllabic to phonemic compression. As in Stone and Moore (2008), the compression ratio was 2.78 for all channels and the compression threshold was 13 dB below the rms level in each channel.

A measure of the effective amount of compression, introduced by Stone and Moore (2004), is the fractional reduction in modulation f_r, which is defined as the proportion of modulation removed by the compressor, plotted as a function of modulation rate. Figure 1 shows f_r functions for the compressors used in this experiment (solid lines) and those used by Stone and Moore (2008) (dashed lines). Compressor speed 3 here led to some reduction in modulation depth (f_r

>0.1) for modulation rates up to about 15 Hz, while compressor speed 1 only reduced the modulation depth for rates up to about 1.8 Hz.

IV. LISTENERS, TRAINING, AND PROCEDURE

Twenty-four native speakers of English [6 male, 18 female, mean age = 20.9 years, standard deviation (SD) = 1.3 years, range = 19–25 years], all university undergraduates or graduates, were recruited. All had audiometric thresholds less than 20 dB HL at the octave frequencies between 125 and 8000 Hz, including the half-octaves of 3000 and 6000 Hz. All attended a 1-h training session and two 2-h testing sessions. Since the entire IEEE corpus was used for testing, and to avoid repeating any of the speech material, the training was based on the Bamford-Kowal-Bench sentence lists (Bench and Bamford, 1979).

The experiment involved 36 different conditions (two values of N_v, three values of N_c for each N_v, three speeds of compression, and two envelope-filter cutoff frequencies). Each condition was assessed using 20 IEEE sentences, each containing five keywords, giving a maximum score of 100 per listener per condition. Data collection was spread evenly over two sessions, each preceded by a small amount of retraining using the adaptive sentence lists (ASL, MacLeod and Summerfield, 1990). The test order of the different conditions was randomized across listeners.

V. RESULTS

Despite the training given, learning effects were apparent in the time-ordered data, perhaps because the materials used for training were not as complex as the IEEE sentences used for testing. To compensate for these effects, quadratic polynomials were fitted to the logarithm of the mean scores as a function of time order (without regard to condition), and the inverse of the fitted mean was used to correct the score of each data point in time order.

The mean time-corrected data, expressed as percent correct are shown in Table I and are illustrated as contour plots in Fig. 2. Each panel in Fig. 2 shows contours of equal performance, plotted as a function of N_c and compressor speed, for one combination of N_v (8 or 16) and the type of envelope filter (E or P). The lightest area in each panel indicates the highest score for the given combination, and the contours are spaced at intervals corresponding to a 2.25% change in intelligibility, which is roughly the smallest statistically significant change. Intelligibility decreased when the number and/or speed of the compressor channels increased. Between the bottom-left of each panel (best performance) and its opposite corner, means scores decreased by about 13% points, except for condition 16P, where the decrease was only about 7% points.

A within-subjects analysis of variance was performed on the time-corrected and arcsine-transformed data, separately for each of the four subgroups of data (8E, 8P, 16E, and 16P), since the TBR varied across subgroups. The factors were N_c and compressor speed. There were significant main effects of both factors ($p<0.001$) for subgroups 8E, 8P, and 16E. For subgroup 16P, the main effects were significant, but

TABLE I. Mean word intelligibility scores in percent correct and SDs (in parentheses), as a function of N_c and speed of compression, grouped by N_v and envelope-filter cutoff frequency.

	Compressor speed		
	1	2	3
$N_v=8$, E filter			
$N_c=4$	65.1 (9.0)	57.1 (11.7)	51.8 (11.6)
$N_c=2$	65.5 (9.7)	63.5 (10.7)	53.9 (10.6)
$N_c=1$	66.7 (9.7)	62.5 (9.6)	60.2 (8.2)
$N_v=8$, P filter			
$N_c=4$	65.3 (11.4)	59.8 (11.8)	53.5 (10.2)
$N_c=2$	68.1 (9.3)	64.8 (9.7)	58.7 (8.0)
$N_c=1$	67.8 (9.9)	63.7 (9.3)	63.8 (10.9)
$N_v=16$, E filter			
$N_c=8$	68.5 (10.9)	64.5 (8.8)	57.4 (12.1)
$N_c=4$	69.7 (10.0)	66.6 (10.2)	60.4 (10.0)
$N_c=2$	70.6 (9.4)	71.4 (10.2)	64.7 (7.2)
$N_v=16$, P filter			
$N_c=8$	63.3 (9.5)	58.5 (12.3)	56.3 (10.8)
$N_c=4$	63.1 (9.3)	61.0 (11.6)	60.0 (12.3)
$N_c=2$	63.3 (10.7)	63.9 (9.7)	61.7 (11.0)

p values were higher at $p=0.016$ and $p=0.02$ for N_c and compressor speed, respectively. The interaction between N_c and compressor speed was significant only for subgroups 8E and 8P ($p=0.035$ and $p=0.019$, respectively).

In summary, the results show that intelligibility decreased when either the number or speed of the compressor channels increased. However, this effect was markedly reduced when high-rate envelope cues were preserved in many channels.

VI. DISCUSSION AND CONCLUSIONS

The data reported here are qualitatively similar to those reported by Stone and Moore (2008) using a noise vocoder preserving only low-rate envelope cues. In both studies, in-

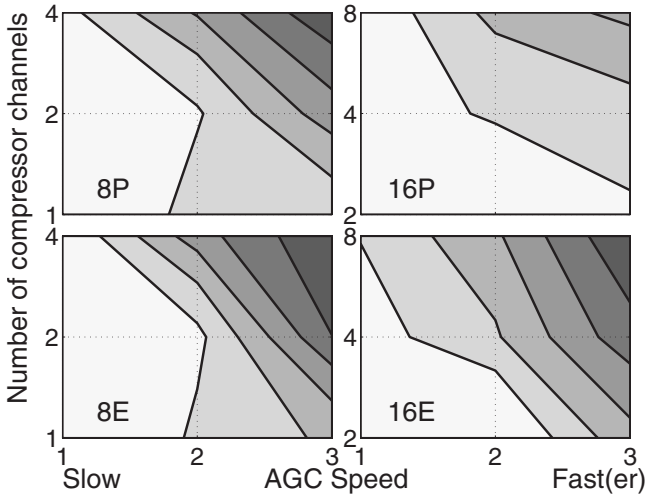


FIG. 2. Contour plots of results shown in Table I separated according to N_v and type of envelope filter. The lightest area indicates good performance, while worsening performance is indicated by the darker areas. The contours are spaced at intervals of 2.25% points.

telligibility decreased as the number of compressor channels and compressor speed increased. However, the rate of decrease with increase in either factor was small when high-rate envelope cues were preserved in a large number of channels (condition 16P here). The pattern of results may be explained in the following way. Low-rate envelope cues are of major importance for speech intelligibility. When those cues are disrupted by multi-channel fast-acting compression, the loss of information can be partially compensated for by the use of higher-rate envelope cues. However, those cues provide less information about articulatory movement than the low-rate cues, and they need to be present in many channels to be reasonably effective.

Since in cochlear implantees the effective number of channels is about 11 (Friesen *et al.*, 2001), it seems likely that fast-acting compression would have deleterious effects on speech intelligibility for implantees. This has been confirmed in a recent study using single-channel compression (Boyle *et al.*, 2009), but remains to be tested using multi-channel compression. For hearing-impaired people using hearing aids, spectral and temporal resolution would usually be sufficient to preserve information comparable to or exceeding that in condition 16P; this would be expected to confer some resistance to the effects of the spectro-temporal degradation produced by fast-acting multi-channel compression. However, even in condition 16P, the compression did have significant deleterious effects on performance.

Stone and Moore (2008) analyzed their results in terms of the trade-off between number of compressor channels and compressor speed needed to maintain a constant level of performance. They reported that, when the number of compressor channels was doubled, the compressor speed needed to be reduced by a factor of about 0.63 for $N_v=12$ or 18. The factors obtained here were slightly higher for conditions 8E, 8P, and 16E, being 0.75, 0.70, and 0.79, respectively, probably because of the higher compression speeds used here. However, the factor for condition 16P at 0.62 was distinctly smaller than for the other conditions. One implication of this is that if a designer of a compression system wishes to maximize the audibility of low-level portions of the signal by using fast compression, then, to avoid reduction in intelligibility when background sounds are present, this needs to be combined with a small number of compression channels; however, more compression channels can be used when the combination of system and user allows high-rate envelope information to be used with high spectral resolution.

ACKNOWLEDGMENTS

This work was supported by the Medical Research Council (UK). One of the authors (C.F.) was supported by a EU Marie Curie Fellowship and a Junior Research Fellowship at Wolfson College, Cambridge, UK.

Bacon, S. P., and Viemeister, N. F. (**1985**). "Temporal modulation transfer functions in normal-hearing and hearing-impaired subjects," Audiology **24**, 117–134.

Baer, T., and Moore, B. C. J. (**1994**). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," J. Acoust. Soc. Am. **95**, 2277–2280.

Bench, J., and Bamford, J. (**1979**). *Speech-Hearing Tests and the Spoken Language of Hearing-Impaired Children* (Academic, London).

Boothroyd, A., Springer, N., Smith, L., and Schulman, J. (**1988**). "Amplitude compression and profound hearing loss," J. Speech Hear. Res. **31**, 362–376.

Boyle, P. J., Büchner, A., Stone, M. A., Lenarz, T., and Moore, B. C. J. (**2009**). "Comparison of dual-time-constant and fast-acting AGC systems in cochlear implants," Int. J. Audiol. **48**, 211–221.

Braida, L. D., Durlach, N. I., De Gennaro, S. V., Peterson, P. M., and Bustamante, D. K. (**1982**). "Review of recent research on multiband amplitude compression for the hearing impaired," in *The Vanderbilt Hearing-Aid Report*, edited by G. A. Studebaker and F. H. Bess (Monographs in Contemporary Audiology, Upper Darby, PA).

Chi, T., Gao, Y., Guyton, M. C., Ru, P., and Shamma, S. (**1999**). "Spectro-temporal modulation transfer functions and speech intelligibility," J. Acoust. Soc. Am. **106**, 2719–2731.

De Gennaro, S., Braida, L. D., and Durlach, N. I. (**1986**). "Multichannel syllabic compression for severely impaired listeners," J. Rehabil. Res. Dev. **23**, 17–24.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

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

Hopkins, K., and Moore, B. C. J. (**2007**). "Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information," J. Acoust. Soc. Am. **122**, 1055–1068.

IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 225–246.

MacLeod, A., and Summerfield, Q. (**1990**). "A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use," Br. J. Audiol. **24**, 29–43.

Moore, B. C. J. (**2007**). *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues*, 2nd ed. (Wiley, Chichester).

Moore, B. C. J. (**2008**). "The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people," J. Assoc. Res. Otolaryngol. **9**, 399–406.

Moore, B. C. J., and Glasberg, B. R. (**2001**). "Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 1067–1073.

Moore, B. C. J., Johnson, J. S., Clark, T. M., and Pluvinage, V. (**1992**). "Evaluation of a dual-channel full dynamic range compression system for people with sensorineural hearing loss," Ear Hear. **13**, 349–370.

Moore, B. C. J., Peters, R. W., and Stone, M. A. (**1999**). "Benefits of linear amplification and multi-channel compression for speech comprehension in backgrounds with spectral and temporal dips," J. Acoust. Soc. Am. **105**, 400–411.

Plomp, R. (**1988**). "The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function," J. Acoust. Soc. Am. **83**, 2322–2327.

Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.

Shannon, R. V. (**1992**). "Temporal modulation transfer functions in patients with cochlear implants," J. Acoust. Soc. Am. **91**, 2156–2164.

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

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

Stone, M. A., and Moore, B. C. J. (**2004**). "Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task," J. Acoust. Soc. Am. **116**, 2311–2323.

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

Whitmal, N. A., Poissant, S. F., Freyman, R. L., and Helfer, K. S. (**2007**). "Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience," J. Acoust. Soc. Am. **122**, 2376–2388.

Yund, E. W., and Buckles, K. M. (**1995**). "Enhanced speech perception at low signal-to-noise ratios with multichannel compression hearing aids," J. Acoust. Soc. Am. **97**, 1224–1240.

Analysis of categorical response data: Use logistic regression rather than endpoint-difference scores or discriminant analysis (L)

Geoffrey Stewart Morrison[a)]
School of Language Studies, Australian National University, Canberra, Australian Capital Territory 0200, Australia

Maria V. Kondaurova
Department of Otolaryngology - Head and Neck Surgery, Indiana University School of Medicine, 699 West Drive - RR044, Indianapolis, Indiana 46202

(Received 17 March 2009; revised 6 August 2009; accepted 11 August 2009)

Example of a typical second-language (L2) speech perception experiment: Synthetic vowel stimuli from a two-dimensional grid of points in which acoustic properties vary systematically in duration and spectral properties are classified as English /i/ or /ɪ/ by L2-English listeners. In a number of studies, the data from such experiments have been analyzed using endpoint-difference scores or discriminant analysis. The current letter describes theoretical problems inherent in the first procedure in general, and in the application of the second procedure to data of this type in particular. Logistic regression is proposed as an alternative, which does not suffer from these problems.
 [DOI: 10.1121/1.3216917]

PACS number(s): 43.71.An, 43.71.Hw, 43.71.Es [AJ] Pages: 2159–2162

I. INTRODUCTION

In cross- and second-language (L2) speech perception research, a common experimental design involves having listeners classify synthetic speech stimuli, where the acoustic properties of the stimuli systematically vary in equal steps along one or more dimensions. The set of stimuli is often referred to as a continuum, although it is actually a grid of equally-spaced points. At least three methods for quantifying the resulting data are attested in the L2 literature: endpoint-difference scores[1–4] (also known as *reliance metrics* or *effect scores*), discriminant analysis,[3,5] and logistic regression.[3,6–8] In the present letter, we will explain that there are theoretical problems with endpoint-difference scores in general, and with discriminant analysis when applied to categorical response data in particular. We recommend logistic regression as a procedure that does not suffer from these problems and briefly demonstrate how it can be used to analyze speech perception data.

For concreteness, we will illustrate our arguments using data derived from an L2 speech perception experiment in a recent paper.[3] Experiment 1: First-language (L1) English listeners, L1-Spanish L2-English listeners, and L1-Russian L2-English listeners classified vowels from an English /i/-/ɪ/ continuum. The continuum consisted of a 99-point two-dimensional grid of linear-predictive-coding resynthesized /bVt/ tokens, in which the vowel duration ranged from 35 to 275 ms in 30 ms steps, and the vowel formant center frequencies ranged from F1=458 Hz, F2=1876 Hz, and F3 =2523 Hz to F1=326 Hz, F2=2056 Hz, and F3 =2943 Hz, respectively, in equal mel steps. The grid was labeled using reference numbers where duration and spectral values of 1 refer to the most /i/-like properties (long duration, and low F1 and high F2) and duration and spectral values of 9 refer to the most /ɪ/-like properties (short duration, and high F1 and low F2). The 81 stimuli were each presented ten times in random order and on each trial the listener classified the stimulus as either English /bit/ or /bɪt/ (there were a total of 16 L1-English listeners, 18 L1-Spanish listeners, and 19 L1-Russian listeners). The data and logistic regression software are available on the first author's website (http://geoff-morrison.net).

II. PROBLEMS WITH END POINT-DIFFERENCE SCORES

In Ref. 3, in Experiment 1, the duration endpoint-difference scores were calculated as the proportion of /ɪ/ responses for all stimuli with duration value of 9 minus the proportion of /ɪ/ responses for all stimuli with duration value of 1. Likewise the spectral endpoint-difference scores were calculated as the difference in the proportion of /ɪ/ responses at the two spectral extremes of the stimulus set.

Figure 1 represents perception data from four L1-Spanish listeners, which are the proportion of /ɪ/ responses at each duration value, pooled over all spectral values. Figure 1 indicates that when classifying English /i/ and /ɪ/, listener 05 made more use of duration cues than listener 17, who in turn made more use of duration cues than listener 11, who in turn made more use of duration cues than listener 10. However, the duration endpoint-difference scores were 0.99, 1, 0.71, and 0.33 respectively. The first two values demonstrate two problems with endpoint-difference scores: *ceiling effect* and *susceptibility to noise*.

A. Ceiling effect

The duration endpoint-difference scores for listeners 05 and 17 are almost the same. As previously pointed out in

a)Author to whom correspondence should be addressed. Electronic mail: geoff.morrison@anu.edu.au

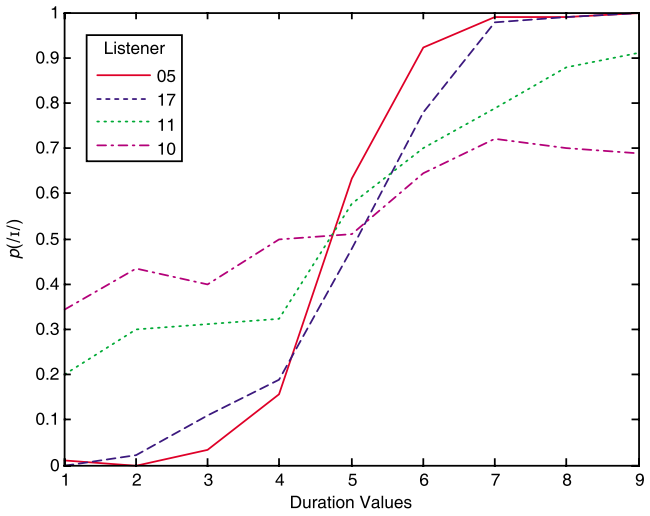


FIG. 1. (Color online) Proportion /ɪ/ responses pooled over all spectral values from selected L1-Spanish listeners. Duration value of 1=275 ms and duration value of 9=35 ms.

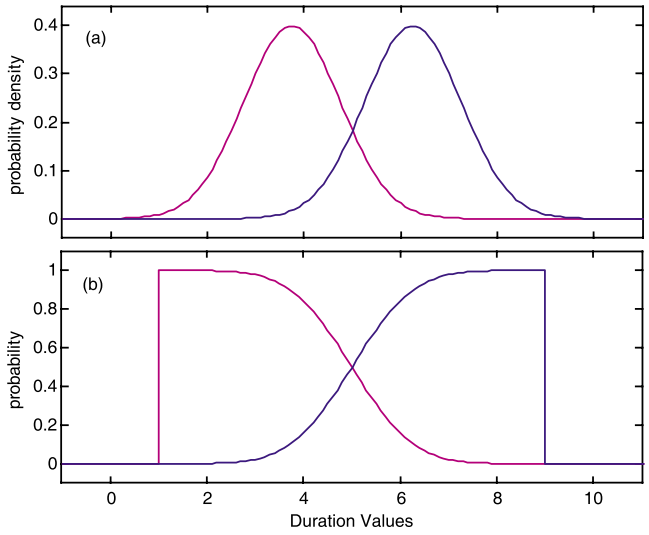


FIG. 2. (Color online) (a) Example of univariate probability density functions from two categories with normal distributions (artificial data) and (b) examples of the distributions of binomial response data from a unidimensional speech classification experiment (artificial data).

Ref. 9, there is a ceiling effect inherent in endpoint-difference scores: If two listeners both give zero /ɪ/ responses at duration value of 1 and 100% /ɪ/ responses at duration value of 9, then they will both have a duration endpoint-difference score of 1 even if over the intermediate duration values the sigmoidal curve of the proportion of /ɪ/ responses is steeper for one listener than for the other. An ideal metric would not suffer from this ceiling effect, and the value given to the steeper response curve of listener 05 would be greater than the value given to the shallower response curve of listener 17.

B. Susceptibility to noise

Since endpoint-difference scores are based on a small subset of the data collected, they are unnecessarily prone to variability due to noise. Although the response curve of listener 05 is steeper than that of listener 17, listener 05 has a lower endpoint-difference score (0.99 versus 1) because out of 90 responses to stimuli with duration value of 1, she gave one /ɪ/ response. An ideal metric would fit a smoothed function to all of the available data and the value given to the steeper response curve of listener 05 would be greater than the value given to the shallower response curve of listener 17. Although we have demonstrated susceptibility to noise with data at or near ceiling, it is a general problem independent of the ceiling effect problem. Ignoring most of the data collected also constitutes a waste of statistical power and a waste of participants' time.

III. PROBLEMS WITH THE APPLICATION OF DISCRIMINANT ANALYSIS TO SPEECH PERCEPTION DATA

Discriminant analysis is a common statistical procedure, descriptions of which can be found in standard textbooks.[10,11] It avoids the ceiling effect inherent in endpoint-difference scores, it fits functions to all the available data, thus reducing problems related to noise, and unlike endpoint-difference scores it allows one to determine the boundary location and allows for multivariate analysis. However, it is not appropriate to apply discriminant analysis to categorical response data because it assumes that the data from each group have a normal distribution (multivariate normal if there is more than one dimension). Figure 2(a) provides a one-dimensional example of two categories with normal distributions, and Fig. 2(b) provides a one-dimensional example of categorical response data (since no response data are collected beyond the range of stimuli tested, the response values below duration 1 and above duration 9 are zero). The distributions of the latter data are clearly not normal, and because of this the means and variances that the discriminant analysis estimates will not be valid. In addition, the mean and covariance estimates will be influenced by the range of values in the stimulus space—if the response curve of a listener happens to be symmetrical around the center of the stimulus space, then the results may not be too bad, but otherwise they will be skewed. Figure 3 represents the effect on a linear discriminant analysis of adding three additional points to a nine-point unidimensional stimulus set. Assuming that the additional points are in a portion of the stimulus space where the listener's responses will be 100% /ɪ/, and that there is no range effect, the listener's response curve will not change, and ideally the fitted curve should not change either; however, the means estimated by a discriminant analysis are further apart and the estimated variance is increased, and hence the fitted posterior probability curve is displaced to the right and has a shallower slope.

IV. LOGISTIC REGRESSION ANALYSIS

A procedure that has the same advantages as discriminant analysis but avoids the problems described above is logistic regression. Logistic regression is a standard statistical procedure, which has been applied in numerous speech perception studies.[12–15] A tutorial on the use of logistic regression with the specific type of perception data discussed here is presented in Ref. 16. The reader is encouraged to

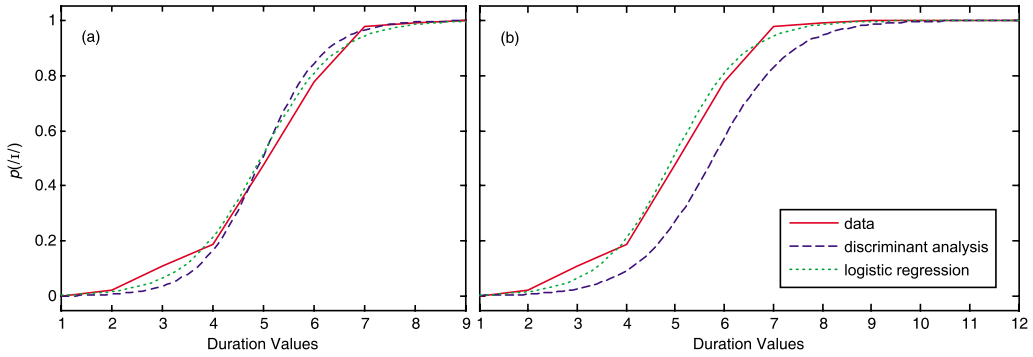


FIG. 3. (Color online) Fits of linear discriminant analysis models and logistic regression models to the proportion of /ɪ/ responses pooled over all spectral values (data from L1-Spanish listener 17). (a) Original data and (b) original data plus three additional duration values with 100% /ɪ/ responses.

refer to Ref. 16 and the works cited therein for a more in-depth coverage than can be presented in the space available here.

Logistic regression is similar to linear regression in that it fits a model with a bias (intercept) and stimulus-tuned (slope) coefficients, but differs in that the model is fitted in a logged odds space. The logged odds transformation converts proportions in the range 0 to 1 into logits in the range $-\infty$ to $+\infty$. Logit values from the fitted model can be converted to probabilities so that lines, planes, and hyperplanes in the logged odds space become sigmoidal curves, surfaces, and hypersurfaces in the probability space. A logistic regression model was fitted to each listener's proportion of /ɪ/ responses. The models included a bias coefficient (α), and duration- and spectrally-tuned coefficients (β_{dur} and β_{spec}) tuned by the duration and spectral properties of the stimuli (x_{dur} and x_{spec}), see Eq. (1):

$$\ln\left(\frac{p(/\text{ɪ}/|x_{\text{dur}},x_{\text{spec}})}{1-p(/\text{ɪ}/|x_{\text{dur}},x_{\text{spec}})}\right)=\alpha+\beta_{\text{dur}}\times x_{\text{dur}}+\beta_{\text{spec}}\times x_{\text{spec}}. \quad (1)$$

The values of the stimulus-tuned coefficients can be used as measures of the perceptual weight of their respective cues, the bias coefficients are also necessary if one wishes to calculate boundary locations. Figure 4 provides a scatterplot of the stimulus-tuned coefficient values, and Fig. 5 provides example probability surface plots from models fitted to the same listeners as in Fig. 1. In contrast to the duration endpoint-difference scores of 0.99, 1, 0.71, and 0.33 for L1-Spanish listeners 05, 17, 11, and 10, respectively, the duration-tuned logistic regression coefficient (β_{dur}) values were 2.01, 1.64, 0.59, and 0.34. The first two β_{dur} values are clearly not subject to the ceiling effect, which made the first two endpoint-difference scores almost identical. Since logistic regression does not assume normal distributions, it is not subject to the same drawbacks as discriminant analysis when applied to this type of data, and unlike the discriminant analysis curve, the logistic regression curves in Figs. 3(a) and 3(b) are almost identical.

Morrison[6,8] proposed and tested a hypothetical (indirect) developmental path for L1-Spanish listeners learning the English /i/-/ɪ/ contrast. For ease of qualitative description, the path can be split into stages: Stage $\frac{1}{2}$, they distinguish English /i/ and /ɪ/ via a category-goodness-assimilation to Spanish /i/, labeling good matches for Spanish /i/ (short stimuli with low F1 and high F2) as English /ɪ/, and labeling poorer matches for Spanish /i/ (longer stimuli with higher F1 and lower F2) as English /i/. Stage 1, perception is duration-based with longer stimuli labeled as English /i/. Stages 2 and 3, the use of spectral cues increases and the use of duration cues decreases approximating L1-English listeners' perception pattern. The results of the logistic regression analyses shown in Fig. 4 (of data independently collected by Kondaurova and Francis[3]) are consistent with Morrison's hypothesized indirect developmental path; in particular, a number of L1-Spanish listeners have the Stage $\frac{1}{2}$ pattern of small negative β_{spec} and small positive β_{dur} values. Interestingly, the distribution of the L1-Russian listeners' perceptual patterns appears to be similar to that of the L1-Spanish listeners, suggesting that the same hypothesized developmental path may apply to L1-Russian listeners learning the English /i/-/ɪ/ contrast.

V. CONCLUSION

We have demonstrated that there are theoretical problems with endpoint-difference scores in general and with discriminant analysis when applied to categorical response data.

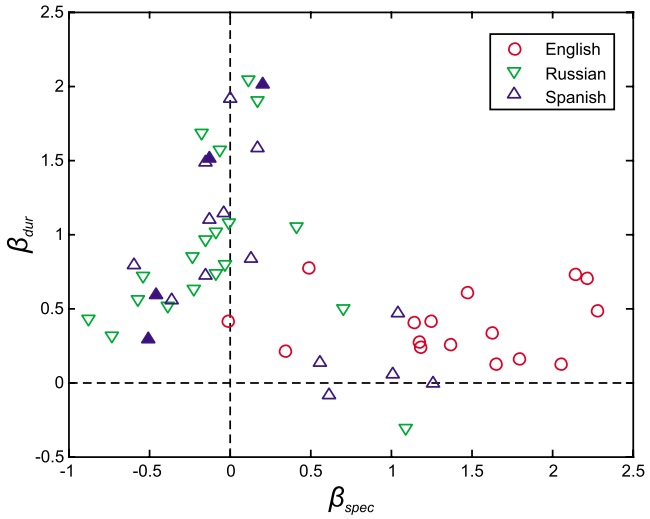


FIG. 4. (Color online) Scatterplot of stimulus-tuned coefficient values from logistic regression models fitted to each listener's response data. The filled symbols correspond to the sample probability surface plots given in Fig. 5.

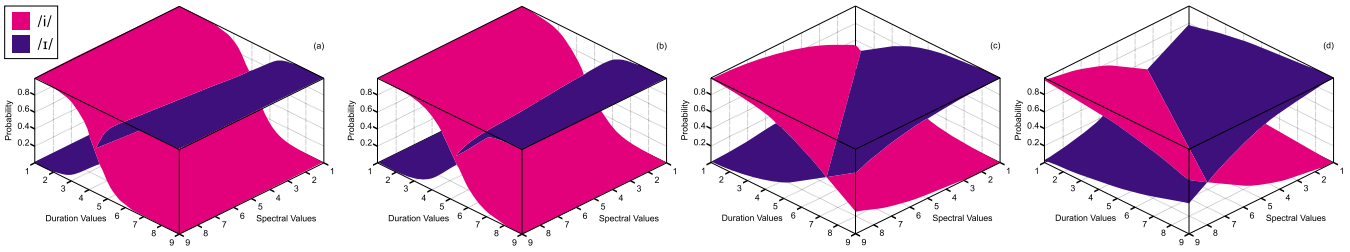


FIG. 5. (Color online) Sample probability surface plots from logistic regression models fitted to the perception data from L1-Spanish listeners (a) 05, (b) 17, (c) 11, and (d) 10.

We have proposed logistic regression as an alternative that does not suffer from these problems and have provided an example of a logistic regression analysis of speech perception data.

[1]P. Escudero and P. Boersma, "Bridging the gap between L2 speech perception research and phonological theory," Stud. Second Lang. Acquis. **26**, 551–585 (2004).

[2]J. E. Flege, O.-S. Bohn, and S. Jang, "Effects of experience on non-native speakers' production of English vowels," J. Phonetics **25**, 437–470 (1997).

[3]M. V. Kondaurova and A. L. Francis, "The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners," J. Acoust. Soc. Am. **124**, 3959–3971 (2008).

[4]X. Wang, "Mandarin listeners' perception of English vowels: Problems and strategies," Can. Acoust. **34**, 15–26 (2006).

[5]G. S. Morrison, "Perception of English /i/ and /ɪ/ by Japanese and Spanish listeners: Longitudinal results," in *Proceedings of the North West Linguistics Conference 2002*, edited by G. S. Morrison and L. Zsoldos (Simon Fraser University Linguistics Graduate Student Association, Burnaby, BC, Canada, 2002), pp. 29–48.

[6]G. S. Morrison, "L1-Spanish speakers' acquisition of the English /i/–/ɪ/ contrast: Duration-based perception is not the initial developmental stage," Lang Speech **51**, 285–315 (2008).

[7]G. S. Morrison, "Perception of synthetic vowels by monolingual Canadian-English, Mexican-Spanish, and Peninsular-Spanish listeners," Can. Acoust. **36**, 17–23 (2008).

[8]G. S. Morrison, "L1-Spanish speakers' acquisition of the English /i/-/ɪ/ contrast II: Perception of vowel inherent spectral change," Lang Speech **52**, 437–462 (2009).

[9]G. S. Morrison, "An appropriate metric for cue weighting in L2 speech perception: Response to Escudero & Boersma (2004)," Stud. Second Lang. Acquis. **27**, 597–606 (2005).

[10]R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. (Wiley, New York, 2000).

[11]T. Hastie, R. Tibshirani, and F. Jerome, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, New York, 2001).

[12]J. R. Benkí, "Place of articulation and first formant transition pattern both affect perception of voicing in English," J. Phonetics **29**, 1–22 (2001).

[13]M. Goudbeek, A. Cutler, and R. Smits, "Supervised and unsupervised learning of multidimensionally varying non-native speech categories," Speech Commun. **50**, 109–125 (2008).

[14]T. M. Nearey, "Speech perception as pattern recognition," J. Acoust. Soc. Am. **101**, 3241–3254 (1997).

[15]R. Smits, J. Sereno, and A. Jongman, "Categorization of sounds," J. Exp. Psychol. Hum. Percept. Perform. **32**, 733–754 (2006).

[16]G. S. Morrison, "Logistic regression modelling for first- and second-language perception data," in *Segmental and Prosodic Issues in Romance Phonology*, edited by P. Prieto, J. Mascaró, and M.-J. Solé (John Benjamins, Amsterdam, 2007), pp. 219–236.

Near resonant bubble acoustic cross-section corrections, including examples from oceanography, volcanology, and biomedical ultrasound

Michael A. Ainslie
Sonar Department, TNO, P.O. Box 96864, 2509 JG The Hague, The Netherlands and Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom

Timothy G. Leighton
Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom

(Received 27 June 2008; revised 23 June 2009; accepted 23 June 2009)

The scattering cross-section σ_s of a gas bubble of equilibrium radius R_0 in liquid can be written in the form $\sigma_s = 4\pi R_0^2/[(\omega_1^2/\omega^2 - 1)^2 + \delta^2]$, where ω is the excitation frequency, ω_1 is the resonance frequency, and δ is a frequency-dependent dimensionless damping coefficient. A persistent discrepancy in the frequency dependence of the contribution to δ from radiation damping, denoted δ_{rad}, is identified and resolved, as follows. Wildt's [*Physics of Sound in the Sea* (Washington, DC, 1946), Chap. 28] pioneering derivation predicts a linear dependence of δ_{rad} on frequency, a result which Medwin [Ultrasonics **15**, 7–13 (1977)] reproduces using a different method. Weston [Underwater Acoustics, NATO Advanced Study Institute Series Vol. **II**, 55–88 (1967)], using ostensibly the same method as Wildt, predicts the opposite relationship, i.e., that δ_{rad} is *inversely* proportional to frequency. Weston's version of the derivation of the scattering cross-section is shown here to be the correct one, thus resolving the discrepancy. Further, a correction to Weston's model is derived that amounts to a shift in the resonance frequency. A new, corrected, expression for the extinction cross-section is also derived. The magnitudes of the corrections are illustrated using examples from oceanography, volcanology, planetary acoustics, neutron spallation, and biomedical ultrasound. The corrections become significant when the bulk modulus of the gas is not negligible relative to that of the surrounding liquid.
 [DOI: 10.1121/1.3180130]

PACS number(s): 43.20.Ks, 43.30.Pc, 43.35.Bf, 43.35.Zc [ADP] Pages: 2163–2175

I. INTRODUCTION

Gas bubbles play an important role in the generation, scattering, and absorption of sound in a liquid.[1] Applications include performance prediction for search sonar or underwater telemetry, acoustical oceanography, medical and industrial ultrasound, and volcanology. The acoustic properties of bubbles are generally well understood, to the extent that acoustical measurements are sometimes used to determine characteristics of bubble clouds such as their size distribution.[2–5] Such acoustical characterization of bubble properties requires a firm foundation in theory.

The purpose of this article is to highlight and resolve a discrepancy that exists at the heart of the currently accepted theory of bubble acoustics. The scattering cross-section, σ_s, of a small spherical bubble of equilibrium radius R_0, undergoing forced linear pulsations at angular frequency ω, is commonly written in the form

$$\sigma_s = \frac{4\pi R_0^2}{(\omega_1^2/\omega^2 - 1)^2 + \delta^2}, \tag{1}$$

where ω_1 is the bubble's pulsation resonance frequency and δ is a dimensionless frequency-dependent parameter known variously as the loss factor, damping constant, or damping coefficient. The term "damping coefficient" is adopted here throughout. The value of δ at resonance is equal to the reciprocal of the Q-factor.

In Sec. II two models for δ are described that differ in the frequency dependence of the damping due to acoustic re-radiation in the free field[6] (known as "radiation damping"). In one of these, published by Wildt[7] and Medwin,[2,8] the radiation damping coefficient is directly proportional to frequency, whereas in the other, published by Andreeva[9] and Weston,[10] the proportionality is an inverse one. This discrepancy has hitherto gone largely unnoticed, to the extent that the authors know of only three publications that mention it.[11–13]

In Sec. III two different derivations for σ_s are provided, with particular attention to establishing the correct frequency dependence of δ for small bubbles. It is shown that the discrepancy is caused in part by ambiguity in the *definition* of δ, and three alternative (though not equivalent) definitions for this parameter are suggested, which can be expressed in terms of the unambiguous damping factor β. The first of the two derivations, which includes thermal damping using a generalization of Weston's method, leads to Eq. (25), including a correction term that is not present in Weston's original formulation. The second, starting from Prosperetti's[14] equation of motion, leads to Eq. (43), of identical form to Eq. (25), and also including the new correction term. This second

derivation, which permits a more general damping factor, leads further to new expressions for the resonance frequency [Eq. (47)] and extinction cross-section [Eq. (67)]. The resulting expression for σ_s is compared with a reference solution due to Anderson,[15] which, unlike the other models considered, has no restriction on bubble size, and serves as ground truth for the situation involving only acoustic radiation losses. In Sec. IV the persistence of the discrepancy and its consequences for the extinction cross-section are discussed. (The Andreeva–Weston model, shown in Sec. III to be the correct one, was last used in the open literature, to the best of the authors' knowledge, more than 40 years ago,[10,11] whereas the incorrect formulation is widely promulgated through standard reviews.[1,16–18]) This is followed in Sec. V by a description of scenarios in which the magnitude of the required correction is not negligible, and conclusions are summarized in Sec. VI.

II. SCATTERING CROSS-SECTION: PUBLISHED RESULTS

In this section some previously published results for the scattering cross-section of a single bubble are considered, stripping them of all forms of damping other than radiation damping. Thus, except where stated otherwise, the effects of viscosity (of the liquid) and thermal conduction (in the gas) are neglected. Surface tension at the boundary is also neglected. These assumptions are made for simplicity and clarity, in order to highlight the discrepancy in the radiation damping term. The publications form two groups, each adopting a different model for the frequency dependence of the radiation damping term.

A. Wildt–Medwin (WM) model

The first relevant publication is the volume edited by Wildt,[7] Chap. 28 of which presents, for the first time, a detailed description of the response of a bubble to ensonification through resonance. Reference 7 offers a clear physical insight into the important physical mechanisms that give rise to damping at and around the resonance frequency.

Wildt's derivation suggests that if acoustic re-radiation is the only loss mechanism, then δ has a linear dependence on frequency. Specifically, Wildt's formula for the radiation damping coefficient is

$$\delta_{\mathrm{WM}}(\omega) = \frac{R_0}{c}\omega, \tag{2}$$

where c is the speed of sound in the surrounding liquid. Equation (2) is used by Medwin[8] to describe the frequency dependence of radiation damping in Eq. (1), and further promulgated by its use in landmark papers[12,19,20] and standard reviews.[1,16–18] The use of $\delta_{\mathrm{WM}}(\omega)$ from Eq. (2) in place of the radiation damping coefficient in Eq. (1) is referred to as constituting the WM model.

B. Andreeva–Weston (AW) model

A form of the damping coefficient that is less well known is derived by Weston[10] and appears for the first time

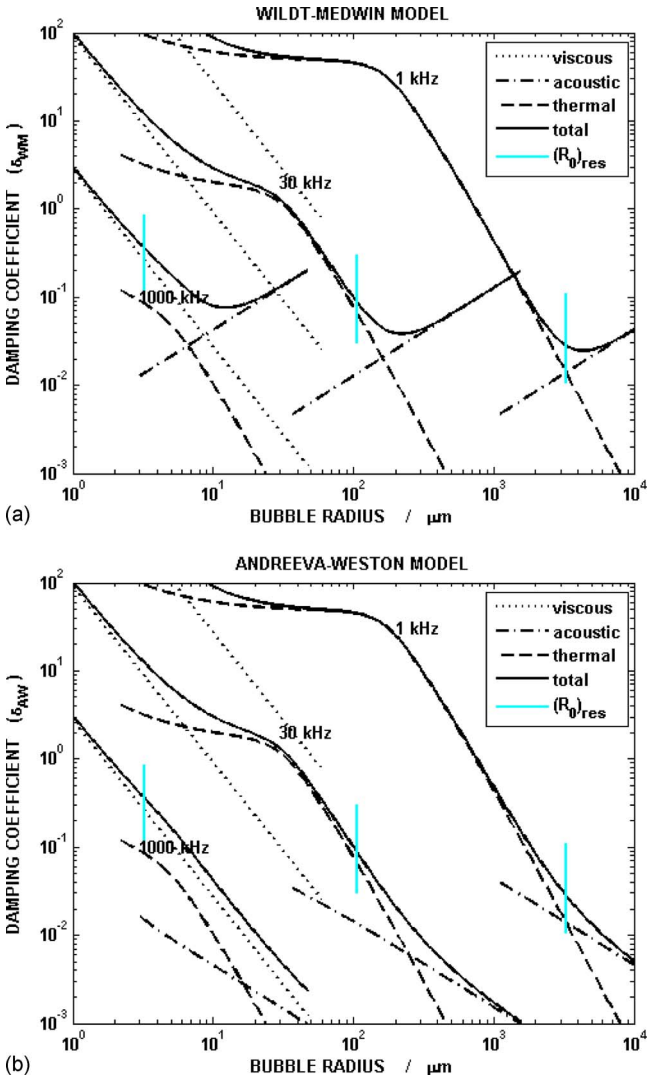


FIG. 1. (Color online) Theoretical damping coefficient vs equilibrium bubble radius for air bubbles in water at atmospheric pressure (0.1 MPa) for acoustic frequencies 1, 30, and 1000 kHz, calculated using (a) the Wildt-Medwin model and (b) the Andreeva–Weston model. The total damping is the black solid line and the contribution due to acoustic radiation is shown as a dash-dotted line (-.-). The remaining curves are for viscous and thermal damping, as indicated by the legend. The resonant bubble radius [denoted $(R_0)_{\mathrm{res}}$] is marked for each frequency using a vertical gray line (cyan online). This quantity is calculated here as the value of the equilibrium bubble radius $R_0=(R_0)_{\mathrm{res}}$ that satisfies the condition $\omega_0(R_0,\omega)=\omega$, where ω_0 is given by Eq. (23) [generalized to incorporate surface tension effects (Ref. 14)].

(without derivation) in the work of Andreeva.[9] Although Weston does not introduce the variable δ explicitly, his derivation of σ_s results in an expression that is consistent with Eq. (1) only if the radiation damping coefficient is *inversely* proportional to frequency, that is, if δ is replaced with δ_{AW}, given by

$$\delta_{\mathrm{AW}}(\omega) = \frac{{\omega_1}^2 R_0}{c}\omega^{-1}. \tag{3}$$

The use of $\delta_{\mathrm{AW}}(\omega)$ from Eq. (3) in place of the radiation damping coefficient in Eq. (1) is referred to as constituting the AW model. A graph similar to Fig. 1 from Medwin's paper,[8] showing the variation in the total damping coefficient with bubble radius R_0 at three frequencies, is presented here

as Fig. 1(a), using Eq. (2) for the radiation damping coefficient, proportional to R_0. Figure 1(b) shows the damping coefficient plotted in the same way, except that the contribution from acoustic radiation is calculated using Eq. (3). Both graphs include viscous and thermal damping, as well as the effects of surface tension. The discrepancy between WM and AW models of radiation damping is apparent from the lower right portion of each solid curve, where the radiation damping term is dominant.

Apart from by Weston himself,[11,21] use of the AW model is rare. The only other publication the authors know of is that of Anderson and Hampton,[12] which presents three different equations for σ_s: first, their Eq. (54) (attributed by Anderson and Hampton[12] to Spitzer[22] and abbreviated here as AH-54), which Anderson and Hampton state is not valid away from resonance, and is identical to the WM model; second, a corrected version, AH-55, which is identical to the AW model; and third, AH-56, which includes additional effects due to viscous and thermal damping and forms the basis of their subsequent calculations of bubble attenuation vs frequency. Once stripped of these extra complications, AH-56 reduces not to AW but to WM, making AH-55 and AH-56 mutually inconsistent.

III. SCATTERING CROSS-SECTION: THEORY

Spherically symmetrical pulsations of a single gas bubble of negligible density in an infinite volume of liquid are considered. Except in Sec. III E the bubble radius is assumed to be small compared with the acoustic wavelength in the liquid. Perturbations to the bubble's radius are assumed small, permitting use of the methods of linear acoustics.

A. Damping factor (β)

Following Morfey,[23] the expression "damping factor" is used in this paper to refer to the parameter β in the equation of motion

$$\ddot{X} + 2\beta\dot{X} + {\omega_{\text{nat}}}^2 X = 0, \tag{4}$$

where the dots represent time derivatives and ω_{nat} is the undamped natural frequency. In the following Eq. (4) is applied to the bubble, with X representing the departure of the bubble's radius (R) from its equilibrium value ($X=R-R_0$). Considering further a sinusoidal forcing term of angular frequency ω, and in general permitting β to vary with frequency, $\beta \rightarrow \beta(\omega)$, the equation of motion then becomes

$$\ddot{R} + 2\beta(\omega)\dot{R} + K(\omega)(R - R_0) = F(\omega)e^{i\omega t}, \tag{5}$$

in which R is understood to be a complex variable and β, K are real parameters that are independent of time, representing resistive and elastic forces, respectively. Like β, the parameters K and F can also be functions of the forcing frequency ω. The effect of inertia is included in the forcing term, so that

$$Fe^{i\omega t} = -\frac{4\pi {R_0}^2}{m_{\text{RF}}^{\text{rad}}} P_F(t), \tag{6}$$

where P_F is the external forcing pressure, $m_{\text{RF}}^{\text{rad}}$ is the radiation mass in the radius-force frame,[1] equal to three times the displaced liquid mass

$$m_{\text{RF}}^{\text{rad}} = 4\pi\rho_0 {R_0}^3, \tag{7}$$

and ρ_0 is the equilibrium density of the liquid.

The particular solution to Eq. (5) is

$$R = R_0 + \frac{F}{K - \omega^2 + 2i\beta\omega} e^{i\omega t}, \tag{8}$$

and hence

$$|R - R_0|^2 = \frac{|F|^2/\omega^4}{(K/\omega^2 - 1)^2 + (2\beta/\omega)^2}. \tag{9}$$

If the term $2\beta/\omega$ is small, then the maximum response occurs when ω is equal to $\sqrt{K}$, which means that K may be approximated in Eq. (9) by the square of the resonance frequency ($K \approx {\omega_1}^2$). If this substitution is made, the similarity between the denominators of Eq. (9) (the radial excursion) and Eq. (1) (the scattering cross-section) makes it tempting to assume further that δ is equal to $2\beta/\omega$, but is it correct to do so? It turns out there is no simple answer to this question, as the true relationship between δ and β depends on the precise definition of δ, which is explored further below.

B. Derivation of the damping coefficient (δ), based on Weston

The following derivation follows the method of Weston,[10] generalized by replacing the specific heat ratio (γ) with a complex polytropic index (denoted Γ). Consider a plane wave p_i of pressure amplitude A and angular frequency ω:

$$p_i = A\exp[i\omega(t - x/c)], \tag{10}$$

incident on a spherical bubble placed with its center at the origin such that the factor F in Eq. (5) is given by

$$F = -\frac{A}{\rho_0 R_0} \tag{11}$$

and

$$P_F = p_i(x = 0) = A\exp(i\omega t). \tag{12}$$

Assume that the scattered wave p_s is a spherical one of amplitude $|B|/r$, such that

$$p_s = (B/r)\exp[i\omega(t - r/c)], \tag{13}$$

where r is the distance from the origin. The scattering cross-section σ_s can then be defined in terms of the ratio B/A as

$$\sigma_s \equiv 4\pi|B/A|^2, \tag{14}$$

Weston's derivation for this ratio is now followed. Euler's equation relates p_s to the radial component of particle velocity u:

$$-\frac{\partial p_s}{\partial r}=\rho_0\frac{\partial u}{\partial t}. \tag{15}$$

Substituting Eq. (13) into Eq. (15) relates the particle velocity in the liquid at distance r from the bubble center to the acoustic pressure associated with the spherical radiated field, $p_s(r)$, giving

$$u(r)=\left(1-\frac{i}{\omega r/c}\right)\frac{p_s(r)}{\rho_0 c}, \tag{16}$$

which, when evaluated at the bubble wall, relates the scattered amplitude to the rate of change of bubble volume V (assuming that departures from the bubble's rest radius are small)

$$\frac{B}{{R_0}^2}(1+i\omega R_0/c)\exp[i\omega(t-R_0/c)]=\frac{i\omega\rho_0}{4\pi{R_0}^2}\frac{\mathrm{d}V}{\mathrm{d}t}. \tag{17}$$

An expression for $\mathrm{d}V/\mathrm{d}t$ can be found by differentiating the polytropic relationship between pressure and volume (PV^Γ =constant, where Γ is the complex polytropic index of the air bubble[17,24]). If the hydrostatic pressure is P_0 and the acoustic pressure inside the bubble is p_b, such that the total interior pressure is P_0+p_b, the result is

$$\mathrm{d}V/\mathrm{d}t=-i\omega p_b V_0/\Gamma P_0, \tag{18}$$

where V_0 is the unperturbed bubble volume. Substituting for $\mathrm{d}V/\mathrm{d}t$ in Eq. (17) and rearranging for the interior acoustic pressure

$$p_b=4\pi\frac{\Gamma P_0 B}{\rho_0 V_0\omega^2}(1+i\omega R_0/c)\exp[i\omega(t-R_0/c)]. \tag{19}$$

In the absence of surface tension, this pressure must equal the sum of the incident and scattered fields [using Eqs. (10) and (13)]:

$$p_b=\left(A+\frac{B}{R_0}e^{-i\omega R_0/c}\right)e^{+i\omega t}. \tag{20}$$

Equating the right hand sides of Eqs. (19) and (20), and solving for the ratio B/A, the result is (introducing the shorthand ε for $\omega R_0/c$)

$$\frac{B}{A}=\frac{R_0}{[(1+i\varepsilon)\Omega^2/\omega^2-1]e^{-i\varepsilon}}, \tag{21}$$

where Ω is a complex parameter given by the equation

$$\Omega(R_0,\omega)^2=4\pi\frac{\Gamma(R_0,\omega)P_0R_0}{\rho_0 V_0(R_0)}, \tag{22}$$

the real part of which is closely related to the pulsation resonance frequency. Specifically, if Γ is real and independent of frequency (which occurs for both isothermal and adiabatic pulsations), then Ω is equal to the bubble's natural frequency. For example, in the adiabatic case Γ and Ω are equal to the specific heat ratio and the Minnaert frequency,[25] respectively.

To proceed further, the following variables are defined

$$\omega_0(R_0,\omega)\equiv\sqrt{\mathrm{Re}[\Omega(R_0,\omega)^2]} \tag{23}$$

and

$$\beta_{\mathrm{th}}(R_0,\omega)\equiv\frac{\mathrm{Im}[\Omega(R_0,\omega)^2]}{2\omega}. \tag{24}$$

The variable ω_1 was introduced earlier as the "resonance frequency" but not properly defined. The new variable ω_0 plays a similar role and can be thought of as a more rigorously defined version of the same variable, though it is more closely related to the natural frequency than the resonance frequency. Wherever the symbol ω_0 is used below, it is always in the sense of Eq. (23).

With these definitions, from Eq. (21) the following generalization of Weston's expression for the scattering cross-section is obtained:

$$\sigma_s=\frac{4\pi{R_0}^2}{\left(\frac{{\omega_0}^2}{\omega^2}-1-2\frac{\beta_{\mathrm{th}}}{\omega}\varepsilon\right)^2+\delta_{\mathrm{AW}}(\omega)^2}, \tag{25}$$

where

$$\delta_{\mathrm{AW}}(\omega)\equiv\mathrm{Im}\frac{p_i(0)}{p_s(R_0)}=2\frac{\beta_{\mathrm{th}}}{\omega}+\frac{{\omega_0}^2}{\omega^2}\varepsilon. \tag{26}$$

If thermal effects are neglected (implying that $\omega_0=\omega_{\mathrm{nat}}$ and $\beta_{\mathrm{th}}=0$), Equation (25) simplifies to the AW model with ω_1 equal to ω_{nat}. The correction term $-2\beta_{\mathrm{th}}\varepsilon/\omega$ in the denominator of Eq. (25) [amounting to a fractional correction to the resonance frequency of $\beta_{\mathrm{th}}(\omega_0)R_0/c$] is new.

Wildt's derivation makes the same assumptions and follows an almost identical procedure as Weston's, so why does it result in a different expression for δ [Eq. (2)]? A close look at Wildt's derivation reveals a subtle error on p. 462. The error occurs in the step from Wi-17 to Wi-22, where the abbreviation Wi-n indicates Eq. (n) from Ref. 7. Specifically, although Wi-13, Wi-16, and Wi-17 are correct to first order in ε, a missing *second* order term is required for the step to Wi-22. To illustrate the nature and importance of this missing term, the expansion $\exp(-i\varepsilon)=1-i\varepsilon-\varepsilon^2/2+O(\varepsilon^3)$ is substituted in Eq. (21) to obtain

$$\frac{B}{A}=\frac{R_0}{\Omega^2/\omega^2-1+\varepsilon^2(\Omega^2/\omega^2+1)/2+i\varepsilon+O(\varepsilon^3)}. \tag{27}$$

Substituting Eq. (27) in Eq. (14) gives Eq. (25) to this order of accuracy, again consistent with the AW model.

One might conclude from this that Wildt's equation for δ [Eq. (2)] is incorrect. Indeed, if δ is defined through Eq. (1), that would seem to be the only *possible* conclusion. In order to be unambiguous, however, such a definition of δ requires the resonance frequency ω_1 to be defined first. Both national[26] and international[27] standards provide different definitions of resonance frequency to choose from, depending on the type of resonance (peak response of, for example, scattered pressure, or bubble wall velocity, or displacement) each leading to a different δ. The most obvious choice for the present purpose would be to define the resonance frequency as the frequency that maximizes σ_s, but this choice leads to an internal contradiction, because this frequency *cannot* be equal to ω_1 in Eq. (1) unless the derivative $\delta'(\omega)$ vanishes at $\omega=\omega_1$. The issue of the resonance frequency is addressed in Sec. III D, but here a second interpretation is considered,

based on Wildt's implied definition (from Wi-32) as the imaginary part of the ratio of R_0 to the scattering amplitude. Denoting this quantity δ_{Wildt}:

$$\delta_{\text{Wildt}}(\omega) \equiv \text{Im}\,\frac{R_0}{B/A}, \tag{28}$$

so that

$$\delta_{\text{Wildt}}(\omega) = \text{Im}\left(\frac{p_i(0)}{p_s(R_0)}e^{-i\varepsilon}\right). \tag{29}$$

With this definition it follows [by substituting Eq. (21) into Eq. (28)] that

$$\delta_{\text{Wildt}}(\omega) = \left(2\frac{\beta_{\text{th}}}{\omega} + \frac{{\omega_0}^2}{\omega^2}\varepsilon\right)\cos\varepsilon - \left(\frac{{\omega_0}^2}{\omega^2} - 1 - 2\frac{\beta_{\text{th}}}{\omega}\varepsilon\right)\sin\varepsilon$$
$$= 2\frac{\beta_{\text{th}}}{\omega} + \varepsilon + \frac{\beta_{\text{th}}}{\omega}\varepsilon^2 - \frac{1}{6}\left(2\frac{{\omega_0}^2}{\omega^2} + 1\right)\varepsilon^3 + O(\varepsilon^4) \tag{30}$$

and

$$\sigma_s = \frac{4\pi {R_0}^2}{\left(\frac{{\omega_0}^2}{\omega^2} - 1\right)^2 + {\delta_{\text{Wildt}}}^2 + \left(\frac{{\omega_0}^4}{\omega^4} - 1\right)\varepsilon^2 + O(\varepsilon^2 {\delta_{\text{Wildt}}}^2)}, \tag{31}$$

replacing Wi-34, which is missing the term $({\omega_0}^4/\omega^4 - 1)\varepsilon^2$.

The physics underlying the source of these discrepancies is illustrated by Eq. (16). The ratio of the local scattered pressure field to the local particle velocity contains both real and imaginary parts [through substitution of Eq. (13) into Eq. (15)]. At the limit of $r \to \infty$, this ratio is real and equal to the impedance of a plane wave, the pressure and velocity are in phase, and indeed locally the wavefront appears planar at $r \to \infty$. At the limit of $r \to 0$, they would be $\pi/2$ out of phase, but this limit cannot be achieved because the bubble wall prevents one tracking back from $r \to \infty$ to $r \to 0$. On such a track the magnitude and phase of the ratio of p_s to u changes from the $r \to \infty$ value, the phase difference increasing towards $\pi/2$ without reaching it. The discrepancy lies in the order to which one approximates by how much the magnitude and phase of the ratio differs from the $r \to 0$ value. This is made clear by the way the $\varepsilon = \omega R_0/c$ terms enter the derivation of Prosperetti:[14] the amplitude term $\varepsilon/(1+\varepsilon^2)^{1/2}$ is retained to second order (equations 3.102, 3.105, and 4.190 of Ref. 1).

It is shown above that, after correcting the error in Wildt's derivation, the results of Wildt[7] and Weston[10] are consistent. It now remains to investigate how Medwin,[2] who uses the damping model of Devin,[28] independently reproduces Wildt's original (uncorrected) result, thus creating a second discrepancy, this time between Weston[10] and Medwin.[2,8] Prosperetti's[14] formulation is now used to address this remaining discrepancy.

C. Alternative derivation of the damping coefficient, based on Prosperetti

Building on the work of Smith,[29] Devin[28] derives an equation of motion for the bubble volume that includes effects of viscous, thermal, and acoustic damping. Prosperetti[14] derives an equation of motion for the bubble radius, including $O(\varepsilon^2)$ correction terms to Devin's equation, that can be written (in the present notation) as

$$\ddot{R} + 2\beta\dot{R} + K(R - R_0) = -\frac{A}{\rho_0 R_0}e^{+i\omega t}, \tag{32}$$

where (neglecting effects of surface tension for consistency with Sec. III B)

$$K = {\omega_0}^2 + \frac{\varepsilon^2}{1+\varepsilon^2}\omega^2, \tag{33}$$

$$\beta = \beta_0 + \frac{\varepsilon}{1+\varepsilon^2}\frac{\omega}{2} \tag{34}$$

and β_0 is the contribution to the damping factor β from mechanisms other than acoustic radiation. [For example, the combined contribution from shear viscosity μ and thermal losses would be $\beta_0 = \beta_{\text{th}} + 2\mu/\rho_0 {R_0}^2$ (Refs. 14 and 30).]

The input impedance Z is defined as the ratio of incident pressure (at $x=0$) to the particle velocity at $r=R_0$:

$$Z \equiv \frac{p_i(x=0)}{u_s(r=R_0)} = i\rho_0 c\varepsilon\left(\frac{K}{\omega^2} - 1 + 2i\frac{\beta}{\omega}\right). \tag{35}$$

An expression for the scattered pressure can then be derived in terms of β by use of Euler's equation (at $r=R_0$):

$$u_s(R_0) = -i\frac{1+i\varepsilon}{\rho_0 c\varepsilon}p_s(R_0). \tag{36}$$

Eliminating u_s from Eqs. (35) and (36) gives

$$\frac{p_i(0)}{p_s(R_0)} = -i\frac{1+i\varepsilon}{\rho_0 c\varepsilon}Z, \tag{37}$$

which can be written as

$$\frac{p_i(0)}{p_s(R_0)} = \frac{{\omega_0}^2}{\omega^2} - 1 - 2\frac{\varepsilon\beta_0}{\omega} + i\left(2\frac{\beta_0}{\omega} + \frac{{\omega_0}^2}{\omega^2}\varepsilon\right). \tag{38}$$

Substitution of Eq. (38) in Eq. (29) results in an equation identical to Eq. (30) except with β_{th} replaced by the more general β_0.

Calculating the scattering cross-section with radiation damping only [i.e., set $\beta_0 = 0$ in the squared modulus of Eq. (38)] the result is Eq. (1), with $\omega_1 = \omega_0$ and Eq. (3) for the damping coefficient, once again in agreement with the AW model, which therefore must be the correct one.

Medwin[2,8] defines δ as

$$\delta_{\text{Medwin}}(\omega) \equiv \frac{2\beta}{\omega}, \tag{39}$$

and then uses this expression for δ in Eq. (1), implicitly (and incorrectly) assuming that it is equal to the imaginary part of the pressure ratio [Eq. (38)]. Substituting Prosperetti's result for β in Eq. (39) gives

$$\delta_{\text{Medwin}}(\omega) = \frac{2\beta_0}{\omega} + \frac{\varepsilon}{1+\varepsilon^2} = 2\frac{\beta_0}{\omega} + \varepsilon - \varepsilon^3 + O(\varepsilon^5). \quad (40)$$

Comparison with the imaginary part of Eq. (38) demonstrates that this assumption results in an unwanted factor $\omega^2/{\omega_0}^2$ in the radiation damping term, coincidentally reproducing Wildt's result for σ_s and reinforcing the erroneous impression that Medwin's and Wildt's expressions for the scattering cross-section are both correct. To lowest order in ε and β_0, the damping coefficients δ_{Wildt} and δ_{Medwin} are equal:

$$\delta_{\text{Medwin}} = \delta_{\text{Wildt}} - \frac{\beta_0}{\omega}\varepsilon^2 - \frac{1}{6}\left(5 - 2\frac{{\omega_0}^2}{\omega^2}\right)\varepsilon^3 + O(\varepsilon^4). \quad (41)$$

A convenient form for σ_s that follows from Eqs. (35) and (37) [with Eq. (39)] is

$$\sigma_s = \frac{4\pi {R_0}^2 (1+\varepsilon^2)^{-1}}{(K/\omega^2 - 1)^2 + {\delta_{\text{Medwin}}}^2}, \quad (42)$$

or (equivalently)

$$\sigma_s = \frac{4\pi {R_0}^2}{\left(\frac{{\omega_0}^2}{\omega^2} - 1 - 2\frac{\beta_0}{\omega}\varepsilon\right)^2 + \left(2\frac{\beta_0}{\omega} + \frac{{\omega_0}^2}{\omega^2}\varepsilon\right)^2}. \quad (43)$$

Equation (43) is derived from Prosperetti's equation of motion. It is identical in functional form to Eq. (25), which is derived using the generalization of Weston's method that was outlined in Sec. III B. The only difference between them is the appearance in Eq. (43) of the more general β_0 instead of β_{th} for the non-acoustic damping factor.

Comparison of Eq. (40) (with $\beta_0 = \beta_{\text{th}}$) with Eq. (26) then gives

$$\delta_{\text{Medwin}} = \delta_{\text{AW}} + \frac{\varepsilon}{1+\varepsilon^2} - \varepsilon\frac{{\omega_0}^2}{\omega^2}. \quad (44)$$

D. Effect of radiation damping on the resonance frequency

The resonance frequency predicted by the AW and MW models and by Eq. (43) are now compared. Specifically, the WM and AW models are considered in the form

$$\sigma_{\text{WM}} = \frac{4\pi {R_0}^2}{\left(\frac{{\omega_0}^2}{\omega^2} - 1\right)^2 + \left(2\frac{\beta_0}{\omega} + \varepsilon\right)^2} \quad (45)$$

and

$$\sigma_{\text{AW}} = \frac{4\pi {R_0}^2}{\left(\frac{{\omega_0}^2}{\omega^2} - 1\right)^2 + \left(2\frac{\beta_0}{\omega} + \frac{{\omega_0}^2}{\omega^2}\varepsilon\right)^2}. \quad (46)$$

In all three cases, the ratio $\sigma_s/{R_0}^2$ is a function of the three parameters ω/ω_0, ε_0, and β_0/ω_0, where $\varepsilon_0 = \omega_0 R_0/c$.

For a pressure resonance, the resonance frequency can be defined[26,27] as the frequency at which the magnitude of the mean square scattered pressure response (proportional to σ_s) is maximized. Adopting this definition and denoting the corresponding resonance frequencies ω_{43}, ω_{WM}, and ω_{AW}, gives the result

$$\left(\frac{\omega_0}{\omega_{43}}\right)^2 = 1 - \frac{2{\beta_0}^2}{{\omega_0}^2} - \frac{{\varepsilon_0}^2}{2}, \quad (47)$$

$$\left(\frac{\omega_0}{\omega_{\text{WM}}}\right)^2 = 1 - \frac{2{\beta_0}^2}{{\omega_0}^2} + \frac{{\varepsilon_0}^2/2}{1 - 2{\beta_0}^2/{\omega_0}^2} + O({\varepsilon_0}^4) \quad ({\varepsilon_0}^2 \ll 1) \quad (48)$$

and

$$\left(\frac{\omega_0}{\omega_{\text{AW}}}\right)^2 = 1 - \frac{2{\beta_0}^2}{{\omega_0}^2} - \frac{{\varepsilon_0}^2}{2} - 2\frac{\varepsilon_0\beta_0}{\omega_0}. \quad (49)$$

No approximation is involved in the derivation of either Eq. (47) [from Eq. (43)] or Eq. (49) [from Eq. (46)], except for the assumption that both β_0 and ω_0 are independent of frequency. That of Eq. (48) [from Eq. (45)] requires the further assumption that ${\varepsilon_0}^2$ is small.

The main point here is that the $O({\varepsilon_0}^2)$ term in Eq. (48) has the wrong sign. This sign error, which can be traced back to the incorrect frequency dependence in the radiation damping term of the WM model, implies that the WM model systematically underestimates the resonance frequency by a fraction of order ${\varepsilon_0}^2$. If a measurement of the resonance frequency were used to estimate the bubble radius, the WM model would lead to a bias of the same order in the inferred radius.

E. Comparison with breathing mode solution for $\beta_0 = 0$

The scattering amplitude for the breathing mode of a spherical gas bubble of arbitrary radius (i.e., without restriction on the magnitude of ε_0) is now evaluated for the case when $\beta_0 = 0$. It has already been shown theoretically that (of the models considered so far) AW [or Eq. (43)] is the correct version. The purpose of the present section is to show that there exists a regime in which ε is large enough for the discrepancy to become an issue, while remaining small enough for the derivation to hold. The amplitude of the scattered wave associated with the pulsating or "breathing" mode (denoted B_{bm}) is determined by[15] (see also Ref. 16)

$$\frac{AR_0}{B_{\text{bm}}(x)} = \frac{\frac{{\omega_0}^2}{\omega^2}\varepsilon\sin\varepsilon - \cos\varepsilon[1 - \xi(x)]}{\frac{{\omega_0}^2}{\omega^2}\cos\varepsilon + \frac{\sin\varepsilon}{\varepsilon}[1 - \xi(x)]} + i\varepsilon, \quad (50)$$

where

$$\xi(x) \equiv \frac{{\omega_0}^2}{\omega^2} + 1 - \frac{3}{x^2}(1 - x\cot x), \quad (51)$$

$$x = \frac{\omega}{\omega_0}\sqrt{\frac{\rho_g}{\rho_0}}, \quad (52)$$

and ρ_g is the equilibrium gas density. The scattering cross-section for the breathing mode is introduced as

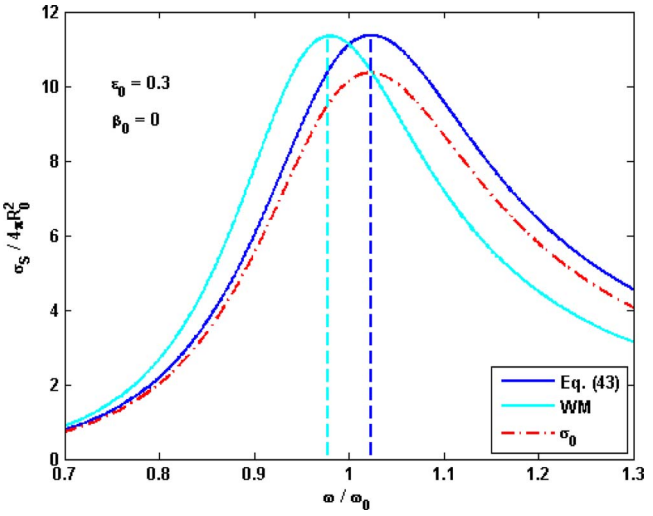


FIG. 2. (Color online) Theoretical pressure response (normalized scattering cross-section, $\sigma_s/4\pi R_0^2$) vs dimensionless frequency ω/ω_0. Black solid line (blue online): Eq. (43); gray line (cyan online): Eq. (45); dash-dot line: Eq. (54) with Eq. (50) and $x=0$. Damping coefficients at resonance are $\varepsilon_0=0.3$ and $\beta_0=0$.

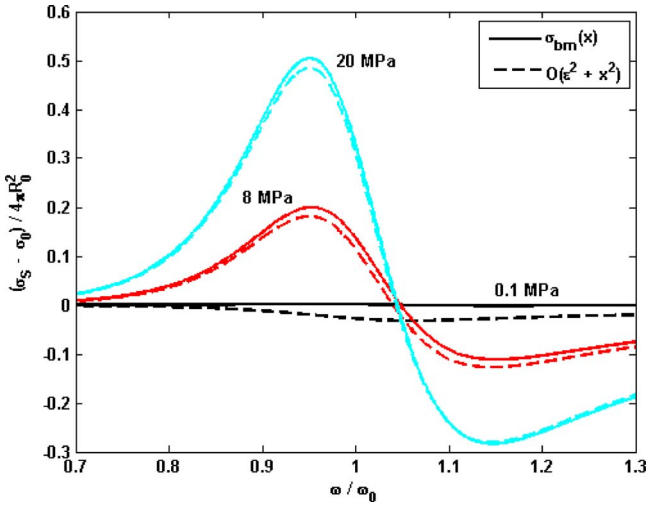


FIG. 3. (Color online) Difference in normalized scattering cross-section $(\sigma_s-\sigma_0)/4\pi R_0^2$ vs dimensionless frequency ω/ω_0, for an air bubble in water at pressure as marked, calculated using Anderson's expression for the breathing mode [Eq. (53), with Eq. (50) for $B_{\rm bm}$, solid lines] and the approximation [Eq. (58), dashed lines]. The corresponding density ratios ρ_g/ρ_0 are 0.00125 (for a pressure of 0.1 MPa), 0.100 (8 MPa), and 0.250 (20 MPa). The radiation damping coefficient at resonance is $\varepsilon_0=0.3$.

$$\sigma_{\rm bm}(x) \equiv 4\pi |B_{\rm bm}(x)/A|^2, \tag{53}$$

with $B_{\rm bm}$ given by Eq. (50), and the variable σ_0 is defined as this cross-section evaluated with zero gas density

$$\sigma_0 \equiv \sigma_{\rm bm}(0). \tag{54}$$

This quantity is plotted in Fig. 2 for the case $\varepsilon_0=0.3$ and compared with the approximations WM and Eq. (43) for the same case, and with $\beta_0=0$, for consistency with Anderson's model [Eq. (43) and AW are identical for this case]. This graph is valid for any gas of negligible density in any liquid. As expected from the sign error in Eq. (48), WM underestimates the resonance frequency (this is caused by the error in radiation damping), whereas Eq. (43) gives the correct resonance frequency. Both WM and Eq. (43) overestimate the maximum breathing mode scattering cross-section. This anomaly is explained below, in the discussion following Eq. (58).

The effect of departures from zero gas density are now considered by plotting the difference between $\sigma_{\rm bm}$ and σ_0 in Fig. 3 (solid curves). This graph shows that an increase in gas density reduces the resonance frequency, an effect that can be understood by considering the behavior of $\sigma_{\rm bm}$ for small values of the ratio ρ_g/ρ_0 as follows. Equation (50) can be approximated, if the gas density is small, using

$$\xi(x) \approx \frac{{\omega_0}^2}{\omega^2} - \frac{x^2}{15}, \tag{55}$$

such that

$$\frac{AR_0}{B_{\rm bm}} \approx \frac{\dfrac{{\omega_0}^2}{\omega^2}\varepsilon \sin\varepsilon - \cos\varepsilon\left(1 - \dfrac{{\omega_0}^2}{\omega^2} + \dfrac{1}{15}\dfrac{\rho_g}{\rho_0}\dfrac{\omega^2}{{\omega_0}^2}\right)}{\dfrac{{\omega_0}^2}{\omega^2}\cos\varepsilon + \dfrac{\sin\varepsilon}{\varepsilon}\left(1 - \dfrac{{\omega_0}^2}{\omega^2} + \dfrac{1}{15}\dfrac{\rho_g}{\rho_0}\dfrac{\omega^2}{{\omega_0}^2}\right)} + i\varepsilon. \tag{56}$$

If ε is also small, expanding to $O(\varepsilon^2)$ gives

$$\frac{AR_0}{B_{\rm bm}} \approx \frac{\dfrac{{\omega_0}^2}{\omega^2} - 1 - \dfrac{1}{15}\dfrac{\rho_g}{\rho_0}\dfrac{\omega^2}{{\omega_0}^2} + \dfrac{\varepsilon^2}{2}\left(1 + \dfrac{{\omega_0}^2}{\omega^2}\right)}{1 + \dfrac{1}{15}\dfrac{\rho_g}{\rho_0}\dfrac{\omega^2}{{\omega_0}^2} - \dfrac{\varepsilon^2}{6}\left(1 + 2\dfrac{{\omega_0}^2}{\omega^2}\right)} + i\varepsilon. \tag{57}$$

Using Eq. (57) for the amplitude of the breathing mode yields the approximate result

$$\sigma_{\rm bm} \approx 4\pi {R_0}^2 \frac{1 - \dfrac{\varepsilon^2}{3}\left(1 + 2\dfrac{{\omega_0}^2}{\omega^2}\right) + \dfrac{2}{15}\dfrac{\rho_g}{\rho_0}\dfrac{\omega^2}{{\omega_0}^2}}{\left(\dfrac{{\omega_0}^2}{\omega^2} - 1 - \dfrac{1}{15}\dfrac{\rho_g}{\rho_0}\dfrac{\omega^2}{{\omega_0}^2}\right)^2 + \varepsilon^2\dfrac{{\omega_0}^4}{\omega^4}}. \tag{58}$$

This expression is evaluated and the difference relative to σ_0 plotted in Fig. 3 (dashed lines) to enable comparison with its counterpart evaluated without these approximations. The graph is calculated for an air bubble in water at a temperature of 283 K. It applies more generally to any gas in any liquid with the density ratios stated in the figure caption. Neglecting the gas density in Eq. (58) results in the AW model for the denominator and an order ε^2 correction in the numerator. This correction to the numerator explains the amplitude anomaly of Fig. 2, without affecting the frequency of resonance.

The resonance frequency associated with Eq. (58) (denoted ω_{58}), to lowest order in ${\varepsilon_0}^2$ and ρ_g/ρ_0, is given by

$$\left(\frac{\omega_0}{\omega_{58}}\right)^2 = 1 - \frac{{\varepsilon_0}^2}{2} + \frac{\rho_g}{15\rho_0}. \tag{59}$$

By a curious numerical coincidence, the two correction terms in Eq. (59) approximately cancel for a bubble of air in water at atmospheric pressure. The precise ratio for an ideal gas at temperature T is

$$\frac{15\varepsilon_0{}^2\rho_0}{2\rho_g} = \frac{45k_BT}{2m_gc^2}\,\mathrm{Re}\,\Gamma, \tag{60}$$

where k_B is Boltzmann's constant (1.381×10^{-23} J/K) and m_g is the average mass of an air molecule (4.82×10^{-26} kg). Using round values of $T=300$ K and $c=1500$ m/s gives approximately 0.81 Re(Γ) for the right hand side, i.e., between 0.8 (for isothermal pulsations) and 1.2 (for adiabatic ones).

It is argued above that the radiation damping can dominate if ε is sufficiently large. This statement seems to conflict with the initial assumption (made in order to satisfy the requirement of uniform pressure at the bubble wall) that ε is "small." Some experimental conditions have forced pragmatic solutions where the tractable approach has been to apply formulations known to be derived assuming that $\varepsilon\to0$, but where this is not the case in practice. The need to ensonify across the range of bubble pulsation resonances to obtain a size distribution for bubbles in a population spanning orders of magnitude, probably meant that Leighton *et al.*[5] worked at up to $\varepsilon\sim0.2$. The application of a two-frequency technique by Newhouse and Shankar[31] probably generated exposure exceeding $\varepsilon\sim2$.

Nevertheless, it is shown above that there exists a regime in which the ε^2 terms are needed, while the derivation remains valid. Prosperetti[14] also argues that $O(\varepsilon^2)$ terms are not only justified, but *necessary* in the regime when $\varepsilon/2\pi$ is small but ε is of order 1. A further counter-argument is one of principle, as follows. A derivation that purports to be accurate to order ε^2 must include *all* terms of that order. Some of the terms might be negligible for some conditions, but the correct way to identify the circumstances in which they may be legitimately neglected is to derive the formally correct solution and only then consider which terms to omit.

F. Confusion caused by the use of dimensionless δ in σ_s

As the preceding text shows, the use of a dimensionless damping coefficient without adequate definition creates confusion. It would not be correct to state categorically that one or other expression for the dimensionless damping coefficient is right or wrong, because any can be perceived as being correct in complying with each respective definition (giving rise, in the present notation, to δ_{Wildt}, δ_{Medwin}, and δ_{AW}). However, the same ambiguity does not apply to the definition of the scattering cross-section, so one *can* make such a statement about σ_s. Thus, the WM model for σ_s [Eq. (45)] is missing a term of order ε^2 in the denominator. This is the same order as δ^2 itself, making it a correction to leading order in the damping term, translating to the sign error in the corresponding correction term in the expression for the resonance frequency ω_{WM} [Eq. (48)]. The confusion can be mitigated by avoiding use of the dimensionless coefficient δ, replacing it with the unambiguous damping factor β as in Eq. (43).

IV. PERSISTENCE OF THE DISCREPANCY, AND THE EXTINCTION CROSS-SECTION

A. Persistence of the discrepancy: The example of ultrasound contrast agents

The result identified as incorrect by the analysis of Sec. III originates from early research related to search sonar[7] and is now in widespread use in acoustical oceanography. (References 18 and 32 are recent examples taken from many possible candidates). A more recent application that is now explored in more detail arises in biomedical acoustics,[33] namely, in the study of ultrasound contrast agents (UCAs), i.e., microbubbles used to enhance the contrast of ultrasound images. This application is chosen because it illustrates how this previously unchallenged discrepancy has been exported to another field and because the case of UCAs provides a convenient demonstration involving the relationship between the extinction and scattering cross-sections. The narrow bubble size range of UCAs promoted a technique of fitting the measured ultrasonic scatter to models of the scattering and extinction cross-sections, in order to produce empirical estimates of, say, the elasticity[34] or frictional losses in the bubble wall[35] in order to determine the mechanical properties of the stabilized bubble wall. Having an incorrect expression for one of the fixed parameters (radiation damping) is unsatisfactory, especially because since its pioneering introduction in the early 1990s,[34–36] the approach became widely used around the world, with the incorrect formulation appearing in dozens of research papers and reviews.[33,37,38] Additional difficulties are exemplified by experiments with UCAs to measure the attenuation of the ultrasonic signal and then compare these data with the computed extinction cross-section for the bubbles. These difficulties are described below.

B. Confusion caused by the use of dimensionless δ in σ_e

The accepted formula for the extinction cross-section (σ_e) of a bubble can be written as[1,2,8,16]

$$\sigma_e = \sigma_s\frac{\delta}{\delta_{\mathrm{rad}}}, \tag{61}$$

where σ_s is given by Eq. (1) and the denominator,

$$\delta_{\mathrm{rad}} = \delta - 2\beta_0/\omega, \tag{62}$$

is the contribution to the damping coefficient from radiation damping alone. Of the various possible definitions for δ though, the following questions are now posed: which one should be used in (a) the right hand side of Eq. (62), (b) the numerator of Eq. (61), and (c) the expression for σ_s [Eq. (1)]?

To answer these questions the definition of the extinction cross-section is considered as the ratio of the mean rate of work done on the bubble to the mean intensity of the incident plane wave. This definition leads to

$$\sigma_e = -\frac{8\pi\rho_0cR_0{}^2}{|p_i|^2}\overline{\mathrm{Re}(p_i)\mathrm{Re}(p_i/Z)}, \tag{63}$$

and hence

$$\sigma_e = \frac{4\pi {R_0}^2}{(K/\omega^2 - 1)^2 + {\delta_{\text{Medwin}}}^2} \frac{\delta_{\text{Medwin}}}{\varepsilon}, \tag{64}$$

which can be written as

$$\sigma_e = \sigma_s \frac{\delta_{\text{Medwin}}}{\varepsilon}(1 + \varepsilon^2). \tag{65}$$

Equation (65) shows that the answer to both (a) and (b) is δ_{Medwin} (or the approximately equivalent δ_{Wildt}), while the correct answer to (c) is shown in Sec. III to be δ_{AW}.

Therefore, if the dimensionless damping coefficient is used to encompass the losses, then one is faced with the unsatisfactory conclusion that, unless correction terms are applied to the currently accepted equations for these cross-sections, there is no single definition of δ that gives the correct result for both σ_s and σ_e, i.e., correct substitution of Eq. (1) into Eq. (61) requires use of both δ_{Medwin} and δ_{AW}:

$$\sigma_e = \frac{4\pi {R_0}^2}{({\omega_1}^2/\omega^2 - 1)^2 + {\delta_{\text{AW}}}^2} \frac{\delta_{\text{Medwin}}}{\delta_{\text{Medwin}} - 2\beta_0/\omega}. \tag{66}$$

The confusion is eliminated by expressing the cross-sections in terms of β_0 (and ε) instead of δ, i.e., using Eq. (43) for the scattering cross-section, with ω_0 given by Eq. (23), and

$$\sigma_e = \sigma_s \frac{2\beta_0/\omega}{\varepsilon}\left(1 + \frac{\omega}{2\beta_0}\varepsilon + \varepsilon^2\right) \tag{67}$$

for the extinction term.

V. EXAMPLE APPLICATIONS

In many circumstances, the contribution from damping due to radiation losses (without which the various scattering models described in Sec. III are in agreement) is small relative to thermal or viscous damping, so the magnitude of any error produced by the choice of an incorrect model is small. The purpose of this section is to discuss the conditions for which the radiation damping might be large enough to cause a significant effect, including numerical examples from a wide range of applications. Since the effect increases as the bulk modulus of the gas becomes no longer insignificant compared to that of the surrounding (possibly bubbly) liquid, these examples cover not only the acoustic monitoring of domestic bubbly products with high void fractions but also of bubbly liquids in extreme conditions (e.g., in coolant, fuel lines, or engineering for deep-ocean and extraterrestrial environments).

A. When are radiation losses large?

The examples considered for Figs. 2 and 3 involve acoustic radiation but no other form of damping. The value of ε^2 at resonance is proportional to the ratio of the bulk modulus of the gas bubble B_{bubble} to that of the surrounding liquid B_{medium}. Because of the relatively low pressure at the sea surface, the underwater acoustics literature is concerned mostly with damping dominated either by thermal conduction (in the case of large bubbles) or viscosity (small ones). In these circumstances the compressibility of the gas bubble is far greater than that of the surrounding liquid, leading to a strong resonance with low radiation loss (small ε). The ratio $B_{\text{bubble}}/B_{\text{medium}}$ will increase if B_{medium} is reduced. For example, if the medium surrounding the bubble in question is itself a bubbly liquid, the medium becomes more compressible than the bubble-free liquid. Examples are found in ship wakes, white caps caused by breaking waves, foams, bubble clouds generated by therapeutic ultrasound, sparging, or as found in the production of metals, pharmaceuticals, foodstuffs, or domestic products, for which void fractions can exceed 1%.[39,40] The ratio will also increase as B_{bubble} increases with increasing static pressure. If the depth of a bubble in the ocean is increased to a few hundred metres, the radiation damping can be dominant, as in the case of a fish bladder,[9] the lung of a deep diving whale or a methane vent at the seabed[41] (noting the additional complication of hydrate formation). Most models of bubble resonance assume that a bubble-free liquid surrounds the bubble in question, and to get large absolute effects (ε^2 of order 0.1) under such circumstances, for a bubble of air in water the static pressure needs to increase to several hundred megapascals, which is not achievable in oceans on earth. Even if such a high static pressure were to exist, the air inside the bubble would liquefy unless the temperature were also increased. For this reason the acoustics of bubbles at high temperature and pressure in a volcano are considered,[42,43] since their presence might influence or indicate eruptions and outgassing hazard.[44–48] Figure 4 illustrates the presence of a high void fraction in a sample of volcanic rock.

FIG. 4. (Color online) Photograph (by TGL) of bubbles in set lava at Timanfaya, Lanzarote. The image is 117 mm wide.

Below some examples are considered. While in some earlier sections of the paper, non-acoustic forms of damping were neglected for clarity, it is important to include these in the quantitative calculations of Sec. V B.

B. Numerical examples including non-acoustic damping ($\beta_0 \neq 0$)

Figure 5 shows σ_s plotted vs frequency for WM, AW, and Eq. (43), using numerical values of $\varepsilon_0 = 0.3$ and $2\beta_0/\omega_0 = 0.3$. Also plotted (vertical dashed lines) are the resonance frequencies associated with each model, as predicted by Eqs. (47)–(49). Figure 5 shows that if the values of ε_0 and β_0 are realized, significant differences arise not just between the WM and AW models but also between both of

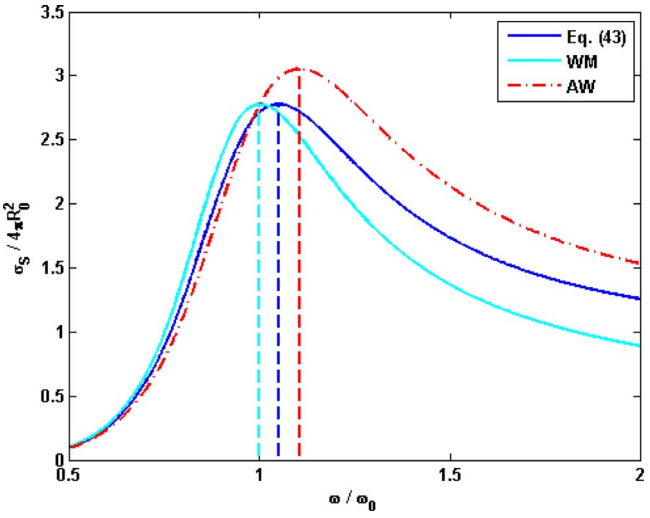


FIG. 5. (Color online) Theoretical pressure response (normalized scattering cross-section) vs dimensionless frequency ω/ω_0 calculated using WM [Eq. (45)], AW [Eq. (46)] and Eq. (43). Damping coefficients are $\varepsilon_0=0.3$ and $2\beta_0/\omega_0=0.3$.

these and Eq. (43). Taking Eq. (47) as a reference on the grounds that the derivation of Eq. (43) makes fewest approximations of the three models considered, it can be seen that neither AW nor WM are wholly accurate: WM underestimates and AW overestimates the resonance frequency [see Eqs. (48) and (49)] by $\varepsilon_0{}^2/2(1-2\beta_0{}^2/\omega_0{}^2)$ and $\varepsilon_0\beta_0/\omega_0$, respectively. The graph applies to any negligible density gas in any liquid.

There are many combinations of temperature and pressure that can give rise to the chosen input values of ε_0 and β_0. To illustrate the diversity, the following generic scenarios are considered:

(1) Case A: air bubbles at atmospheric pressure (e.g., in ship wakes and breaking waves,[49] foodstuff,[39] and cement paste[40]) and the multiphase reactors used in chemical, biochemical, environmental, pharmaceutical, or petrochemical industries.[50]
(2) Case B: methane bubbles in seawater at a depth of order 1000 m (notwithstanding the formation of hydrates at this depth)[41,51,52] (or deep water blowout[53]).
(3) Case C: carbon dioxide bubbles at high temperature and pressure in a volcano.[44–47,54–56]
(4) Case D: nitrogen bubbles in ethane lake on Titan.[57,58]
(5) Case E: helium bubbles in liquid mercury (neutron spallation target).[59–64]

For these gas-liquid mixtures the following notation is introduced (see Table I):

(a) subscript g denotes properties of the gas (e.g., ρ_g for the gas density),
(b) subscript w denotes properties of the host liquid (e.g., c_w for speed of sound in the bubble-free liquid), and
(c) subscript m denotes properties of the gas-liquid mixture (e.g., B_m for its bulk modulus).

Figure 5 is applicable to each of these scenarios. The properties of the gas-liquid mixture are specified by means of the gas bulk modulus B_g, mixture density ρ_m, specific heat ratio γ of the gas, and temperature T. These four parameters may be chosen freely for any given value of ε_0 and β_0/ω_0. Once chosen, the first two of them (B_g and ρ_m) determine the bulk modulus B_m

$$B_m = \frac{3B_g}{\varepsilon_0{}^2} \tag{68}$$

and the sound speed c_m of the mixture (column 6)

$$c_m = \sqrt{\frac{B_m}{\rho_m}}. \tag{69}$$

The bubble radius R_0 can take any value. Once chosen, its value determines the undamped natural frequency (column 7)

$$\omega_0 = R_0{}^{-1}\sqrt{\frac{3B_g}{\rho_m}} \tag{70}$$

and "equivalent" viscosity, which is defined as (column 8)

$$\mu_{\text{eq}} \equiv R_0\sqrt{3B_g\rho_m}\frac{\beta_0}{2\omega_0}. \tag{71}$$

Thus, μ_{eq} is the shear viscosity that would be required, if viscosity were the only non-acoustic damping mechanism, to achieve the non-acoustic damping factor β_0. For example, a bubble radius of 1 mm gives a resonance frequency $(\omega_0/2\pi)$ of between 2 kHz (case C) and 32 kHz (case E) and equivalent viscosity between 1.5 Pa s (case A) and 39 Pa s (case C).

The gas density corresponding to these conditions is shown in the last column [$(\rho_g)_{\text{ad}}$, calculated from B_g assuming adiabatic pulsations]

TABLE I. Defining parameters (in bold) for cases A (air bubbles in wake), B (methane vent), C (carbon dioxide bubbles in molten lava), D (nitrogen bubbles in ethane lake on Titan), and E (helium bubbles in mercury spallation target). Parameters in remaining columns are calculated using the expressions given in the text.

Case	B_g (MPa)	ρ_m (kg m^{-3})	γ	T (K)	c_m (m s^{-1})	$\omega_0 R_0$ (m s^{-1})	$\mu_{\text{eq}}R_0^{-1}$ (Pa s mm^{-1})	$(\rho_g)_{\text{ad}}$ (kg m^{-3})
A	**0.14**	**1 000**	**7/5**	**280**	68.3	20.5	1.54	1.25
B	**10**	**1 000**	**4/3**	**280**	577	173	13.0	51.9
C	**35**	**2 600**	**4/3**	**1470**	670	201	39.2	95.2
D	**0.21**	**630**	**7/5**	**95**	38.9	31.6	1.49	5.36
E	**0.60**	**13 200**	**5/3**	**300**	105	11.7	11.6	0.581

$$(\rho_g)_{\mathrm{ad}} = \frac{m_g B_g}{\gamma k_B T}. \tag{72}$$

Together with the information from the table, this value can be used to estimate the required void fraction using

$$U = \frac{\rho_w - \rho_m}{\rho_w - \rho_g}, \tag{73}$$

provided that the bubble-free liquid density ρ_w is known. To do so, Wood's equation[65] is used in the form (accepting the use of this low-frequency approximation for the purposes of this first-order calculation)

$$\left(\frac{1}{B_g} - \frac{1}{B_m}\right)\rho_w{}^2 + \left(\frac{\rho_g}{B_m} - \frac{\rho_m}{B_g}\right)\rho_w + \frac{\rho_m - \rho_g}{c_w{}^2} = 0. \tag{74}$$

Using this equation it is found that the required void fraction is approximately 3% for all cases considered, even though the static pressure (as given by B_g/γ) varies by more than two orders of magnitude. The physical reason for this result is that, unless the void fraction is very low, increasing the pressure increases the bulk modulus of the gas-liquid mixture surrounding the bubble in question almost as much as it does that of the gas bubble. This can be seen more clearly by writing Eq. (73) in the form

$$U = \frac{\varepsilon_0{}^2/3 - B_g/B_w}{1 - B_g/B_w}. \tag{75}$$

For the cases considered, B_g is small compared with B_w, from which it follows that $U \approx \varepsilon_0{}^2/3 = 0.03$. The difficulty of applying Wood's equation if the bubbles are at or near resonance is recognized,[66,67] but the above simple analysis suggests that large values of ε_0 are unlikely to be achieved by high pressure alone.

VI. SUMMARY AND CONCLUSIONS

The dimensionless damping coefficient δ introduced by Wildt[7] and developed further by Medwin[2] are considered. Particular attention is paid to the role of radiation damping in determining the through-resonance frequency dependence of the scattering cross-section σ_s of a single spherical gas bubble in terms of the parameter ε, defined as the product of acoustic wave number and bubble radius. Specific conclusions are as follows

(1) Published theoretical results of Andreeva[9] and Weston[10] for σ_s are not consistent with those of Wildt and Medwin. The AW model, which has not been used in open literature for more than 40 years, is correct to order ε^2 in the denominator of σ_s. The WM model, which is in widespread use, is missing a term of this order and thus requires a leading order correction to the damping term.
(2) A generalization of Weston's derivation is used to obtain a new expression for σ_s [Eq. (25)], which simplifies to the AW model if non-acoustic damping is neglected. The same equation is then derived by application of Euler's equation to the input impedance obtained from Prosperetti's equation of motion, leading to Eq. (43). This second approach leads also to a new equation for the extinction cross-section σ_e [Eq. (67)].
(3) The magnitude of the effect, as measured by the difference in resonance frequency between the different scattering models, though often small, can become significant in some realistic conditions. It is of order 10% if ε and $2\beta_0/\omega$ are both equal to 0.3 at resonance (and is proportional to ε^2). This requires either a bubble under very high pressure (comparable with the bulk modulus of the surrounding liquid) or a bubble in a highly compressible liquid. Possibilities explored include a ship wake at atmospheric pressure, methane vents under pressure at the seabed, carbon dioxide bubbles in molten lava and helium bubbles in a neutron spallation target. For all cases considered, the required void fraction according to Wood's equation is close to 3% (i.e., $\varepsilon_0{}^2/3$.)
(4) The zeroth order term from Anderson's expansion for the scattering cross-section of a fluid sphere of arbitrary radius and density is simplified and used to confirm the accuracy of the AW model for the case with gas density and non-acoustic damping coefficient both negligible.
(5) Three different definitions of δ are considered, denoted δ_{AW} [defined by Eq. (26)], δ_{Wildt} [Eq. (28)], and δ_{Medwin} [Eq. (39)]. Of these, δ_{AW} is required in Eq. (1) to obtain the correct frequency dependence for σ_s, while either δ_{Wildt} or δ_{Medwin} (and *not* δ_{AW}) must be used in Eq. (65) for σ_e. Unless correction terms are applied to the currently accepted equations for σ_s and σ_e, there is no single definition of δ that gives the correct result for both cross-sections.
(6) In situations for which acoustic radiation is the main form of damping (e.g., in water under high static pressure), the WM model underestimates the resonance frequency. Its use to infer the bubble radius from an acoustical measurement would therefore lead to a systematic bias.

ACKNOWLEDGMENTS

One of the authors (M.A.A.) acknowledges formative discussions with Dr. D. E. Weston concerning the interaction of sound with fish swimbladders. He regrets now that none of these were about the frequency dependence of radiation damping. He also thanks Dr. A. J. Robins for his encouragement to pursue the discrepancy between the WM and AW models of radiation damping. Constructive comments by two anonymous reviewers helped us improve the final manuscript. This work was sponsored in part by the Defence Research and Development Department of the Netherlands Ministry of Defence (M.A.A) and in part by the UK Engineering and Physical Sciences Research Council (EPSRC), the UK Natural Environment Research Council (NERC), the UK Science and Technology Facilities Council (Rutherford Appleton Laboratory), and the Oak Ridge National Laboratories, Tennessee, Spallation Neutron Source (ORNL is managed by UT-Battelle, LLC, under contract DE-AC05-00OR22725 for the U.S. Department of Energy) (T.G.L).

[1]T. G. Leighton, *The Acoustic Bubble* (Academic, London, 1994).
[2]H. Medwin, "Counting bubbles acoustically: A review," Ultrasonics **15**, 7–13 (1977).
[3]A. M. Sutin, S. W. Yoon, E. J. Kim, and I. N. Didenkulov, "Nonlinear acoustic method for bubble density measurements in water," J. Acoust. Soc. Am. **103**, 2377–2384 (1998).
[4]D. V. Farmer, S. Vagle, and D. Booth, "A free-flooding acoustical resonator for measurement of bubble size distributions," J. Atmos. Ocean. Technol. **15**, 1132–1146 (1998).
[5]T. G. Leighton, S. D. Meers, and P. R. White, "Propagation through nonlinear time-dependent bubble clouds, and the estimation of bubble populations from measured acoustic characteristics," Proc. R. Soc. London, Ser. A **460**, 2521–2550 (2004).
[6]T. G. Leighton, P. R. White, C. L. Morfey, J. W. L. Clarke, G. J. Heald, H. A. Dumbrell, and K. R. Holland, "The effect of reverberation on the damping of bubbles," J. Acoust. Soc. Am. **112**, 1366–1376 (2002).
[7]R. Wildt, editor, Chapter 28, "Acoustic theory of bubbles," *Physics of Sound in the Sea*, NDRC Summary Technical Report Div. 6, Vol. 8 (Washington, DC, 1946), pp. 460–477. [Devin (Ref. 28) and Anderson and Hampton (Ref. 12) both refer to an earlier unpublished report by Spitzer (Ref. 22) that Devin credits in part to Willis. The author(s) of Ref. 22 might also have written Wildt's Chap. 28].
[8]H. Medwin, "Acoustical determination of bubble-size spectra," J. Acoust. Soc. Am. **62**, 1041–1044 (1977).
[9]I. B. Andreeva, "Scattering of sound by air bladders of fish in deep sound-scattering ocean layers," Akust. Zh. **10**, 20–24 (1964); English translation in Sov. Phys. Acoust. **10**, 17–20 (1964).
[10]D. E. Weston, "Sound propagation in the presence of bladder fish," *Underwater Acoustics* NATO Advanced Study Institute Series Vol. **II**, Copenhagen, edited by V. M. Albers (Plenum, New York, 1967), pp. 55–88.
[11]D. E. Weston, "Acoustic interaction effects in arrays of small spheres," J. Acoust. Soc. Am. **39**, 316–322 (1966).
[12]A. L. Anderson and L. D. Hampton, "Acoustics of gas-bearing sediments I. Background," J. Acoust. Soc. Am. **67**, 1865–1889 (1980).
[13]M. A. Ainslie and T. G. Leighton, "The influence of radiation damping on through-resonance variation in the scattering cross-section of gas bubbles," *Underwater Acoustics Measurements: Technologies and Results*, Second International Conference and Exhibition, FORTH, Crete, 25–29 June 2007, edited by J. Papadakis and L. Bjørnø, pp. 571–576.
[14]A. Prosperetti, "Thermal effects and damping mechanisms in the forced radial oscillations of gas bubbles in liquids," J. Acoust. Soc. Am. **61**, 17–27 (1977).
[15]V. C. Anderson, "Sound scattering from a fluid sphere," J. Acoust. Soc. Am. **22**, 426–431 (1950).
[16]H. Medwin and C. S. Clay, *Fundamentals of Acoustical Oceanography* (Academic, Boston, 1998).
[17]L. M. Brekhovskikh and Yu. P. Lysanov, *Fundamentals of Ocean Acoustics*, 3rd ed. (Springer-Verlag, New York, 2003).
[18]D. R. Jackson and M. D. Richardson, *High-Frequency Seafloor Acoustics* (Springer, New York, 2007).
[19]R. H. Love, "Resonant acoustic scattering by swimbladder-bearing fish," J. Acoust. Soc. Am. **64**, 571–580 (1978).
[20]K. W. Commander and A. Prosperetti, "Linear pressure waves in bubbly liquids: Comparison between theory and experiments," J. Acoust. Soc. Am. **85**, 732–746 (1989).
[21]D. E. Weston, "Assessment methods for biological scattering and attenuation in ocean acoustics," BAeSEMA Report No. C3305/7/TR-1, Director Science (Sea), Ministry of Defence, London, 1995.
[22]L. Spitzer, Jr., "Acoustic properties of gas bubbles in a liquid," OSRD Report No. 1705 Division of War Research, Columbia University, 1943. This report has not been seen by the present authors.
[23]C. L. Morfey, *Dictionary of Acoustics* (Academic, San Diego, 2001).
[24]C. S. Clay and H. Medwin, *Acoustical Oceanography: Principles and Applications* (Wiley, New York, 1977), p. 463.
[25]M. Minnaert, "On musical air-bubbles and the sounds of running water," Philos. Mag. **16**, 235–248 (1933).
[26]Definition 5.05 resonance frequency, American National Standard Acoustical Terminology, ANSI S1.1-1994 (ASA 111-1994), Revision of ANSI S1.1-1960 (R1976) (Acoustical Society of America, New York, 1994).
[27]Electropedia (IEV online), http://www.electropedia.org/iev/iev.nsf (Last viewed February, 2009).
[28]C. Devin, Jr., "Survey of thermal, radiation, and viscous damping of pulsating air bubbles in water," J. Acoust. Soc. Am. **31**, 1654–1667 (1959).
[29]F. D. Smith, "On the destructive mechanical effects of the gas-bubbles liberated by the passage of intense sound through a liquid," Philos. Mag. **19**, 1147–1151 (1935).
[30]G. Houghton, "Theory of bubble pulsation and cavitation," J. Acoust. Soc. Am. **35**, 1387–1393 (1963).
[31]V. L. Newhouse and P. M. Shankar, "Bubble size measurements using the nonlinear mixing of two frequencies," J. Acoust. Soc. Am. **75**, 1473–1477 (1984).
[32]T. R. Hahn, "Low frequency sound scattering from spherical assemblages of bubbles using effective medium theory," J. Acoust. Soc. Am. **122**, 3252–3267 (2007).
[33]X. Yang and C. C. Church, "A model for the dynamics of gas bubbles in soft tissue," J. Acoust. Soc. Am. **118**, 3595–3606 (2005).
[34]N. de Jong, L. Hoff, T. Skotland, and N. Bom, "Absorption and scatter of encapsulated gas filled microspheres: Theoretical considerations and some measurements," Ultrasonics **30**, 95–103 (1992).
[35]N. de Jong and L. Hoff, "Ultrasound scattering properties of Albunex microspheres," Ultrasonics **31**, 175–181 (1993).
[36]N. de Jong, "Acoustic properties of ultrasound contrast agents," Ph.D. thesis, Erasmus University of Rotterdam, Rotterdam, The Netherlands (1993).
[37]L. Hoff, P. C. Sontum, and B. Hoff, "Acoustic properties of shell-encapsulated, gas-filled ultrasound contrast agents," Proc.-IEEE Ultrason. Symp. **2**, 1441–1444 (1996).
[38]M. Chan, K. Soetanto, and M. Okujima, "Simulations of contrast effects from free microbubbles in relation to their size, concentration and acoustic properties," Jpn. J. Appl. Phys., Part 1 **36**, 3242–3245 (1997).
[39]G. M. Campbell and E. Mougeot, "Creation and characterisation of aerated food products," Trends Food Sci. Technol. **10**, 283–296 (1999).
[40]W. Punurai, J. Jarzynski, J. Qu, K. E. Kurtis, and L. J. Jacobs, "Characterization of entrained air voids in cement paste with scattered ultrasound," NDT & E Int. **39**, 514–524 (2006).
[41]E. Suess, M. E. Torres, G. Bohrmann, R. W. Collier, D. Rickert, C. Goldfinger, P. Linke, A. Heuser, H. Sahling, K. Heeschen, C. Jung, K. Nakamura, J. Greinert, O. Pfannkuche, A. Trehu, G. Klinkhammer, M. J. Whiticar, A. Eisenhauer, B. Teichert, and M. Elvert, "Sea floor methane hydrates at Hydrate Ridge, Cascadia margin," in *Natural Gas Hydrates: Occurrence, Distribution, and Detection*, Geophys. Monograph Series Vol. **124**, edited by C. K. Paull and W. P. Dillon (American Geophysical Union, Washington, DC, 2001), pp. 87–98.
[42]M. J. Buckingham and M. A. Garcés, "A canonical model of volcano acoustics," J. Geophys. Res. **101**, 8129–8151 (1996).
[43]M. A. Garcés, "On the volcanic waveguide," J. Geophys. Res. **102**, 22547–22564 (1997).
[44]S. Vergniolle and C. Jaupart, "Dynamics of degassing at Kilauea Volcano, Hawaii," J. Geophys. Res. **95**, 2793–2809 (1990).
[45]S. Vergniolle and G. Brandeis, "Origin of the sound generated by Strombolian explosions," Geophys. Res. Lett. **21**, 1959–1962 (1994).
[46]S. Vergniolle, G. Brandeis, and J.-C. Mareschal, "Strombolian explosions: 2. Eruption dynamics determined from acoustic measurements," J. Geophys. Res. **101**, 20449–20466 (1996).
[47]S. Vergniolle and J. Caplan-Auerbach, "Acoustic measurements of the 1999 basaltic eruption of Shishaldin volcano, Alaska 2. Precursor to the Subplinian phase," J. Volcanol. Geotherm. Res. **137**, 135–151 (2004).
[48]J. B. Johnson, R. C. Aster, and P. R. Kyle, "Volcanic eruptions observed with infrasound," Geophys. Res. Lett. **31**, L14604 (2004).
[49]E. C. Monahan and N. Q. Lu, "Acoustically relevant bubble assemblages and their dependence on meteorological parameters," IEEE J. Ocean. Eng. **15**, 340–345 (1990).
[50]M. Simonnet, C. Gentric, E. Olmos, and N. Midoux, "Experimental determination of the drag coefficient in a swarm of bubbles," Chem. Eng. Sci. **62**, 858–866 (2007).
[51]A. J. Walton, M. G. Gunn, and G. T. Reynolds, "The quality factor of oscillating bubbles as an indication of gas content with particular reference to methane," IEEE J. Ocean. Eng. **30**, 924–926 (2005).
[52]J. Peckmann and V. Thiel, "Carbon cycling at ancient methane-seeps," Chem. Geol. **205**, 443–467 (2004).
[53]Ø. Johansen, "*DeepBlow*—a Lagrangian plume model for deep water blowouts," Spill Sci. Technol. Bull **6**, 103–111 (2000).
[54]M. A. Garcés, "Theory of acoustic propagation in a multi-phase stratified liquid flowing within an elastic-walled conduit of varying cross-sectional area," J. Volcanol. Geotherm. Res. **101**, 1–17 (2000).
[55]M. A. Garcés, S. R. McNutt, R. A. Hansen, and J. C. Eichelberger, "Application of wave-theoretical seismoacoustic models to the interpretation of explosion and eruption tremor signals radiated by Pavlof volcano,

Alaska," J. Geophys. Res. **105**, 3039–3058 (2000).
[56]O. Navon, A. Chekhmir, and V. Lyakhovsky, "Bubble growth in highly viscous melts: theory, experiments, and autoexplosivity of dome lavas," Earth Planet. Sci. Lett. **160**, 763–776 (1998).
[57]T. G. Leighton and P. R. White, "The sound of Titan: A role for acoustics in space exploration," Acoustics Bulletin **29**, 16–23 (2004).
[58]T. G. Leighton, P. R. White, and D. C. Finfer, "The sounds of seas in space," in Proceedings of the International Conference on Underwater Acoustic Measurements, Technologies and Results, Crete, 2005, edited by J. S. Papadakis and L. Bjørnø, pp. 833–840.
[59]D. Felde, B. Riemer, and M. Wendel, "Development of a gas layer to mitigate cavitation damage in liquid mercury spallation targets," J. Nucl. Mater. **377**, 155–161 (2008).
[60]H. Kogawa, T. Shobu, M. Futakawa, A. Bucheeri, K. Haga, and T. Naoe, "Effect of wettability on bubble formation at gas nozzle under stagnant condition," J. Nucl. Mater. **377**, 189–194 (2008).
[61]T. Lu, R. Samulyak, and J. Glimm, "Direct numerical simulation of bubbly flows and application to cavitation mitigation," J. Fluids Eng. **129**, 595–604 (2007).
[62]B. W. Riemer, P. R. Bingham, F. G. Mariam, and F. E. Merrill, "Measurement of gas bubbles in mercury using proton radiography," in Proceedings of the Eighth International Topical Meeting on Nuclear Applications and Utilization of Accelerators (ACCAPP'07) (2007), pp. 531–537.
[63]K. Okita, S. Takagi, and Y. Matsumoto, "Propagation of pressure waves, caused by thermal shock, in liquid metals containing gas bubbles," J. Fluid Sci. Technol. **3**, 116–128 (2008).
[64]R. Samulyak, T. Lu, and Y. Prykarpatskyy, "Direct and homogeneous numerical approaches to multiphase flows and applications," Lect. Notes Comput. Sci. **3039**, 653–660 (2004).
[65]A. B. Wood, *A Textbook of Sound* (Bell, London, 1946).
[66]F. S. Henyey, "Corrections to Foldy's effective medium theory for propagation in bubble clouds and other collections of very small scatterers," J. Acoust. Soc. Am. **105**, 2149–2154 (1999).
[67]S. G. Kargl, "Effective medium approach to linear acoustics in bubbly liquids," J. Acoust. Soc. Am. **111**, 168–173 (2002).

Measurement of acoustic streaming in a closed-loop traveling wave resonator using laser Doppler velocimetry

Cyril Desjouy,[a] Guillaume Penelet, Pierrick Lotton, and James Blondeau
Laboratoire d'Acoustique de l'Université du Maine, UMR CNRS 6613, Avenida Olivier Messiaen, 72085 Le Mans Cedex 9, France

(Received 25 March 2009; revised 24 August 2009; accepted 25 August 2009)

This paper deals with the measurement of acoustic particle velocity and acoustic streaming velocity in a closed-loop waveguide in which a resonant traveling acoustic wave is sustained by two loudspeakers appropriately controlled in phase and amplitude. An analytical model of the acoustic field and a theoretical estimate of the acoustic streaming are presented. The measurement of acoustic and acoustic streaming velocities is performed using laser Doppler velocimetry. The experimental results obtained show that the curvature of the resonator impacts the acoustic velocity and the profile of acoustic streaming. The quadratic dependence of the acoustic streaming velocity on the acoustic pressure amplitude is verified and the measured cross-sectional average streaming velocity is in good agreement with the value predicted by the theoretical estimate.
 [DOI: 10.1121/1.3238162]

PACS number(s): 43.25.Nm [OAS] Pages: 2176–2183

I. INTRODUCTION

Acoustic streaming is a net mean flow which is generated by sound. This nonlinear effect is known for more than a century,[1] and there is renewed interest in studying this effect because it may be used, for instance, to enhance heat transfer,[2] to generate fluid motion in microfluidic devices,[3,4] or to drive ultrasonic motors.[5,6] Acoustic streaming can also disturb the operation of thermoacoustic engines[7] because it is responsible for generally unwanted heat convection within the device. Even though the influence of acoustic streaming on the efficiency of thermoacoustic engines has been already studied,[8,9] it is still poorly understood. Thus, its effects on the operation of engines are usually almost empirically controlled, using jet pumps[10,11] and/or tapered tubes.[12] The characterization of acoustic streaming in thermoacoustic engines is indeed a difficult task due to large temperature gradients and complicated shapes of the different elements of the engine.

Efforts have been devoted to the theoretical description of acoustic streaming in acoustic resonators[13–16] and in thermoacoustic devices.[17–21] The precise measurement of acoustic streaming velocity has also been recently achieved by Thompson *et al.*[22,23] and Moreau *et al.*[24] using laser Doppler velocimetry (LDV) and appropriate signal processing. Thompson *et al.* measured the Lagrangian outer (i.e., outside acoustic boundary layers) streaming velocity in a standing wave and they studied the influences of the temperature gradients and the fluid inertia on acoustic streaming. They notably demonstrated that the dependence of viscosity on temperature impacts the acoustic streaming velocity as predicted by Rott.[13] They also showed that small temperature gradients induce significant discrepancies between the observed velocities and any available theoretical results. Moreau *et al.* measured both inner/outer and slow/fast streaming velocities using LDV and they focused attention to the spatial distribution of acoustic streaming near the walls (inside the acoustic boundary layers). It is, however, remarkable that, while recent developments of thermoacoustic engines make use of a closed-loop path to increase the efficiency,[10,11,25] most of the studies mentioned above are concerned with standing wave devices. Just a few studies deal with traveling wave devices for which the existence of a closed-loop notably allows the streaming flow to be nonzero across the section of the resonator.

Thus, this paper aims at contributing to a better understanding of acoustic streaming behavior in closed-loop acoustic resonators. The device studied in this paper is not a thermoacoustic engine. It consists of an annular waveguide in which the acoustic field is a resonant acoustic traveling wave generated and controlled by two loudspeakers (Fig. 1). Laser Doppler velocimetry (LDV) measurements of acoustic particle velocity and acoustic streaming velocity are performed in this study, and the results are compared to simplified theoretical models. It should be mentioned that, though the idea of using a closed-loop path to generate a guided traveling wave is not new (see, for instance, Refs. 26–28), we did not find in the literature any report of experiments in the device which is described below.

In Sec. II, the experimental apparatus is briefly described. In Sec. III, an analytical modeling of the acoustic field and a theoretical estimate of the acoustic streaming velocity are presented. The LDV measurement of acoustic particle velocity and acoustic streaming velocity are presented and discussed in Sec. IV.

II. EXPERIMENTAL APPARATUS

A schematic representation of the experimental device is shown in Fig. 1(a). It consists of an annular resonator of unwrapped length $L=2.12$ m and with a square cross-section

[a]Author to whom correspondence should be addressed. Electronic mail: cyril.desjouy@univ-lemans.fr

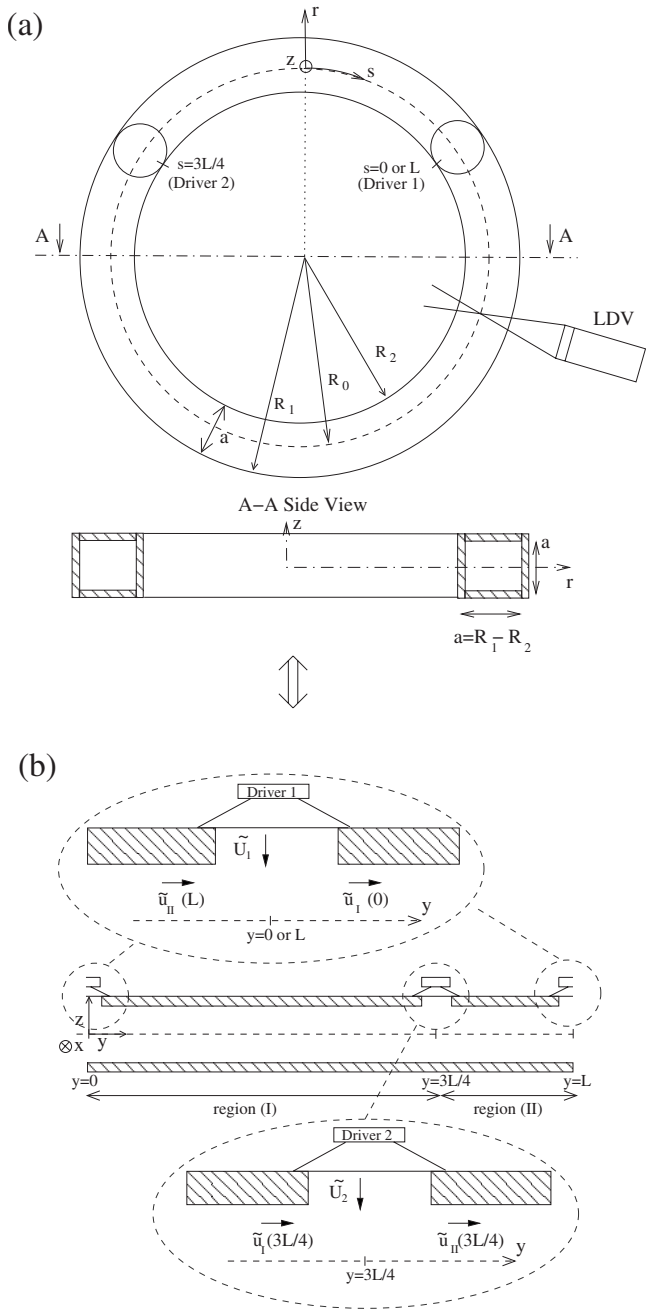


FIG. 1. (a) Schematic representations of the experimental device (top view and side view). (b) Unwrapped schematic representation of the experimental device, the cylindrical system of coordinates (r,θ,z) [or (r,s,z) with $s=\theta L/2\pi$] being replaced by the Cartesian coordinates (x,y,z). The drivers localized at s or $y=0$ (or L) and s or $y=3L/4$ are represented with details on the volume velocities and their orientations. Simultaneous measurements of acoustic and acoustic streaming velocities in the longitudinal direction (s or y) are performed using a commercial laser Doppler velocimeter (Dantec Dynamics, Flowlite 1D).

$S_W=a\times a=7.5\times 7.5$ cm^2. The waveguide, which is filled with air at atmospheric pressure and at room temperature (20 °C), is made of Plexiglass™ to allow LDV measurements of acoustic particle velocity and acoustic streaming velocity. Two electrodynamic drivers (loudspeakers Audax PR170MO), set at positions $s=0$ and $s=3L/4$, and coupled to the waveguide through a circular hole of diameter a, allow to generate an acoustic wave inside the resonator. Particular attention is paid to the accurate control of the drivers: a laser Doppler vibrometer (Polytec OFV 300) is used to measure both amplitude and phase of the oscillating velocities at the center of the loudspeaker membranes. Three microphones (Brüel & Kjær 4136) are flush-mounted along the resonator (located at $s=0.57$ m, $s=0.66$ m, and $s=0.75$ m from the driver 1 set at $s=0$).

The specificity of the experimental device is that, under some circumstances, it is possible to sustain a resonant rotating wave[29] which has the characteristics of a traveling wave in terms of spatial distribution of the acoustic field (and in terms of phase shift between acoustic pressure and velocity fluctuations). More precisely, it has been demonstrated (see, for instance, Ref. 26) that if the drivers are separated by a distance of $L/4$, if each driver is sustaining sound at a frequency f which corresponds to the first natural mode of the air column ($f=c_0/L\approx 161$ Hz, where $c_0\approx 342$ m s^{-1} is the adiabatic speed of sound), and if the drivers displacements are equal in amplitude but $\pi/2$ out of phase, then the resulting acoustic wave propagating into the waveguide behaves like a resonant traveling wave (provided that the dissipation of acoustic energy in the waveguide is negligible). This is due to the fact that each driver generates two counterpropagating acoustic waves in the resonator, two of them propagating in one direction being coherently additive (in direction $+\mathbf{s}$, for instance, if the driver 2 has a phase delay of $-\pi/2$ compared to driver 1) while the two other waves canceling out.

III. ANALYTICAL DESCRIPTION OF THE ACOUSTIC FIELD AND OF THE ACOUSTIC STREAMING

A. Acoustic volume velocity

A simplified analytical description of the acoustic field is presented in this section. It is assumed here that the curvature of the resonator has no influence on the propagation of the acoustic wave inside the resonator, so that the closed-loop waveguide [Fig. 1(a)] is assumed to be equivalent to the unwrapped waveguide [Fig. 1(b)], the "closed-loop" property of the resonator being taken into account by ensuring continuity of acoustic pressure and acoustic velocity between both ends of the unwrapped resonator. Assuming that a plane acoustic wave takes place in the waveguide, the acoustic pressure inside the resonator can be written $p(y,t)=\Re(\tilde{p}(y)e^{-i\omega t})$, where $\Re$ denotes taking the real part of the complex argument and where $\tilde{}$ denotes the complex amplitude. In the following, the axial component of the acoustic particle velocity is denoted $v_y(x,y,z,t)=\Re(\tilde{v}_y(x,y,z)e^{-i\omega t})$. The two loudspeaker membranes provide harmonic motion at frequency $f=\omega/(2\pi)=c_0/L$. The complex amplitudes of acoustic volume velocities generated by the loudspeaker 1, located at $y=0$, and the loudspeaker 2, located at $y=3L/4$, are denoted by $\tilde{U}_1$ and $\tilde{U}_2$, respectively. Two regions, denoted by I and II [Fig. 1(b)], split the closed-loop resonator. The complex amplitude $\tilde{p}_{\mathrm{I,II}}$ of acoustic pressure in each region is written as follows:

$$\tilde{p}_{\mathrm{I,II}}(y)=\tilde{p}^{+}_{\mathrm{I,II}}(y_{\mathrm{I,II}})e^{+ik_w(y-y_{\mathrm{I,II}})}+\tilde{p}^{-}_{\mathrm{I,II}}(y_{\mathrm{I,II}})e^{-ik_w(y-y_{\mathrm{I,II}})}, \quad (1)$$

where $y_{\mathrm{I,II}}$ denote any abscissa in regions I and II, respectively, and where

$$k_w = k_0\sqrt{1+\frac{f_\nu+(\gamma-1)f_\kappa}{1-f_\nu}} \tag{2}$$

is the complex wave number which accounts for viscous and thermal dissipation in the vicinity of the resonator's walls (the functions f_ν and f_κ characterize the viscous and thermal coupling between the wall of the resonator and the oscillating fluid,[7] $k_0=\omega/c_0$, and γ is the specific heat ratio). The complex amplitudes of the acoustic volume velocities $\tilde{u}_{\mathrm{I,II}}=S_W\cdot\langle\tilde{v}_{y_{\mathrm{I,II}}}(x,y,z)\rangle$, where $\langle\cdots\rangle$ is used to denote the cross-sectional spatial average, are obtained from the acoustic pressure as follows:[7]

$$\tilde{u}_{\mathrm{I,II}}(y)=\frac{S_W(1-f_\nu)}{i\omega\rho_0}\frac{\partial\tilde{p}_{\mathrm{I,II}}}{\partial y}, \tag{3}$$

where ρ_0 stands for the mean density of fluid. Setting $y_{\mathrm{I}}=0$ and $y_{\mathrm{II}}=3L/4$, and assuming that, at the loudspeaker locations, there is continuity of acoustic pressure $[\tilde{p}_{\mathrm{I}}(0)=\tilde{p}_{\mathrm{II}}(L)$ and $\tilde{p}_{\mathrm{I}}(3L/4)=\tilde{p}_{\mathrm{II}}(3L/4)]$ and continuity of acoustic volume velocity $[\tilde{u}_{\mathrm{II}}(L)+\tilde{U}_1=\tilde{u}_{\mathrm{I}}(0)$ and $\tilde{u}_{\mathrm{I}}(3L/4)+\tilde{U}_2=\tilde{u}_{\mathrm{II}}(3L/4)]$, the constants $\tilde{p}_{\mathrm{I}}^{\pm}(y_{\mathrm{I}})$ and $\tilde{p}_{\mathrm{II}}^{\pm}(y_{\mathrm{II}})$ can be expressed as functions of $\tilde{U}_{1,2}$. In particular, if $\tilde{U}_2=\tilde{U}_1e^{i\phi}$, this leads to the following expressions for the acoustic volume velocity in the entire resonator:

$$\tilde{u}_{\mathrm{I}}=\frac{i\tilde{U}_{\mathrm{I}}}{4\sin(k_wL/2)}[(e^{-i(k_w(L/4)-\phi)}+e^{-ik_w(L/2)})e^{ik_wy} - (e^{i(k_w(L/4)+\phi)}+e^{ik_w(L/2)})e^{-ik_wy}], \tag{4}$$

$$\tilde{u}_{\mathrm{II}}=\frac{i\tilde{U}_1}{4\sin(k_wL/2)}[(e^{ik_w(L/4)}+e^{-i(k_w(L/2)-\phi)})e^{ik_w(y-(3L/4))} - (e^{-ik_w(L/4)}+e^{i(k_w(L/2)+\phi)})e^{-ik_w(y-(3L/4))}]. \tag{5}$$

Moreover, notifying that for the given angular frequency $\omega=2\pi c_0/L$, the viscous and thermal boundary layer thicknesses $\delta_\nu=\sqrt{2\nu/\omega}$ and $\delta_\kappa=\sqrt{2\kappa/\omega}$ (ν and κ being kinematic viscosity and thermal diffusivity of fluid, respectively) are small compared to the hydraulic radius a of the resonator, the viscous and thermal functions $f_{\nu,\kappa}$ can be approximated by[30]

$$f_{\nu,\kappa}\overset{\delta_{\nu,\kappa}\ll a}{\approx}(1+i)\frac{\delta_{\nu,\kappa}}{a}. \tag{6}$$

Due to this, the complex wavenumber is approximated by $k_w\approx k_0(1+\epsilon)+ik_0\epsilon$, where ϵ is a small parameter given by $\epsilon=\delta_\kappa[(\gamma-1)+\sqrt{\sigma}]/(2a)\ll 1$ and where $\sigma=\nu/\kappa$ denotes the Prandtl number. Tuning adequately the relative phase shift ϕ between the membranes displacements of the loudspeakers allows to generate a given kind of acoustic wave. In particular, if $\phi=-\pi/2$, the approximate expressions of the acoustic volume velocities $\tilde{u}_{\mathrm{I,II}}$ obtained by expanding Eqs. (4) and (5) over the small parameter ϵ, can be written as follows:

$$\tilde{u}_{\mathrm{I}}(y)=\frac{-\tilde{U}_1}{4\epsilon\pi(1-i)}\{-2e^{ik_0y}-\epsilon g_1(y)+O(\epsilon^2)\}, \tag{7}$$

$$\tilde{u}_{\mathrm{II}}(y)=\frac{-\tilde{U}_1}{4\epsilon\pi(1-i)}\left\{2ie^{ik_0(y-(3L/4))}+\epsilon g_2\left(y-\frac{3L}{4}\right)+O(\epsilon^2)\right\}, \tag{8}$$

where

$$g_1(\xi)=\pi(1-i)\left[\cos(k_0\xi)+\left(1-\frac{4\xi}{L}\right)e^{ik_0\xi}\right], \tag{9}$$

$$g_2(\xi)=\pi(1+i)\left[3\cos(k_0\xi)-\left(1+\frac{4\xi}{L}\right)e^{ik_0\xi}\right]. \tag{10}$$

This result clearly shows that under the conditions mentioned above, the acoustic volume velocity can be separated into two components: the first one (of order of magnitude $1/\epsilon$) corresponds to a resonant acoustic wave traveling along the direction $+y$ and the second one (of order 1) corresponds to some spatial variations in acoustic amplitude due to the viscous and thermal boundary layers effects near the walls. So, as long as the parameter ϵ can be considered as small, the present device can be used as a traveling wave resonator.

B. Cross-sectional average streaming velocity

When a high level resonant acoustic traveling wave is sustained into the resonator, an acoustic streaming flow is generated by nonlinear effects, which presents a nonzero cross-sectional average value due to the closed-loop geometry. An approximate expression of this streaming velocity can be obtained using a successive approximation approach, assuming (1) that the boundary layer approximation is valid, (2) that the effect of the curvature of the resonator is neglected, and (3) that the dependence of kinematic viscosity ν of fluid on temperature T_0 is taken into account (namely,[13] $\nu\propto T_0^\beta$, with $\beta=0.73$).

In the following, the subscript m is used to denote the second order in magnitude mean flow generated by acoustic streaming. In order to compare the results of measurements with the theoretical results, the cross-sectional average value of streaming velocity $\langle v_{ym}(y)\rangle$ is expressed as a function of acoustic pressure amplitude when the waveguide has a circular cross-section of diameter a. The details of calculation, which are based on previous works,[18] are given in Appendix. The average value along the closed-loop waveguide $1/L\oint\langle v_{ym}\rangle\cdot dy$ of the cross-sectional average streaming velocity $\langle v_{ym}\rangle$ is found to be proportional to acoustic intensity as follows:

$$\frac{1}{L}\oint\langle v_{ym}\rangle\cdot dy=\alpha p_{\mathrm{rms}}^2, \tag{11}$$

where p_{rms} is the root-mean-square of the acoustic pressure amplitude inside the waveguide, and where the parameter $\alpha\approx-1.8\times10^{-7}$ m s^{-1} Pa^{-2}. It should be mentioned that acoustic streaming velocity is not constant along the waveguide, but the predicted variations in $\langle v_{ym}\rangle$ are lower than 0.5% of its average value. It is also interesting to note that the direction of acoustic streaming is opposite to the direction of the traveling wave.

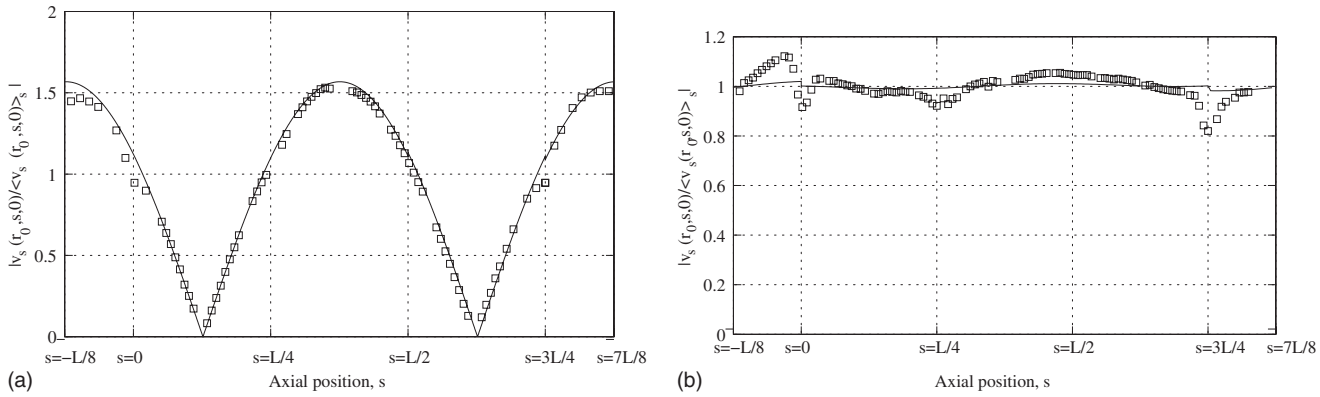


FIG. 2. Normalized distributions of the axial velocity fields $|v_s(R_0,s,0)/\langle v_s(R_0,s,0)\rangle_s|$ as function of the axial position s: squares correspond to the experiment and full lines correspond to the theory. (a) $\phi = 0$; (b) $\phi=-\pi/2$.

IV. MEASUREMENT PROCEDURE

A. Hardware and signal processing

For all the experimental results which are presented in the following, the absolute values of the amplitudes of the displacements of the loudspeaker membranes are equal ($|\tilde{U}_1|=|\tilde{U}_2|$), and the operating frequency $f\approx(2\pi c_0)/L \approx 161$ Hz corresponds to the frequency of the first natural acoustic mode of the air column. We have verified experimentally that this frequency also corresponds to a resonance of the complete device (the presence of the loudspeakers do not affect the acoustic resonance of the closed-loop waveguide). A commercial single component laser Doppler velocimeter (Dantec Dynamics, 158.4 mm standard front lens) is used to measure velocity field. The dimensions of the fringe volume are of about 50 μm in the direction of measurement (along the s-coordinate) by 410 μm wide (along the r-coordinate) and 50 μm high (along the z-coordinate). The LDV probe can move in the transverse direction r [or x if the unwrapped geometry of Fig. 1(b) is considered]. The waveguide can rotate around its center, allowing to make measurements along the axial direction s (or y). The signal processing which is used here is similar to that used by Moreau *et al.*[24] The signal is captured by a photomultiplier and is analyzed by the BSA system (Dantec Dynamics 57N20-BSA burst spectrum analyzer). Acquisition are performed over 40 000 samples in order to reach convergence of the measurement results. Then, given a phase reference, a postprocessing algorithm allows to bring back all samples on a single acoustic period. A least squares method is used to estimate the Eulerian particle velocity. The average value (over an acoustic period) of the least squares fit of the signal corresponds to the streaming velocity v_{sm} (or v_{ym} in case of unwrapped waveguide), while the oscillating component corresponds to the acoustic particle velocity $v_s(t)$ [or $v_y(t)$]. The measurement accuracy is provided by the signal to noise ratio (SNR) defined as[31] $\mathrm{SNR}=10\log(V_s^2/2\mathcal{V}[n(t)])$, where V_s denotes the peak amplitude of the acoustic particle velocity $v_s(t)$ and where $\mathcal{V}[n(t)]$ is the variance of the noise $n(t)$ (which is the difference between measured and estimated acoustic velocities). This SNR is used to calculate the minimal errors attributed to the data acquisition and the signal processing system, and in the following, LDV measurements are considered as valid for a SNR higher than 20 dB. However, it is worth noting that there are additional uncertainties which are due to the experimental device itself. There are indeed some parameters like the evolution of room temperature or the accuracy of the probe positioning, which we tried to monitor with care. These parameters are not taken into account in the definition of the SNR, but may impact the accuracy of the measurements significantly. During measurements, we took care that the variations in room temperature do not exceed 1 K, and we considered that the accuracy of probe positioning is lower than 1 mm.

It is also worth noting that, in our device, the cascade process of higher harmonics generation should be considered because it distorts the waveform and because harmonics generation may contribute to the generation of acoustic streaming (this contribution is neglected in Sec. III B). However, for the largest amplitude of acoustic velocities used (1.9 m/s), the amplitude of the second harmonic of particle velocity does not exceed 5.5% of the amplitude of the first harmonic. The assumption of a monofrequency acoustic field is thus expected to be valid because the magnitude of the acoustic streaming generated by this second-harmonic is expected to be less than 0.3% of the magnitude of the streaming generated by the first harmonic.

B. Acoustic particle velocity measurements

The spatial distribution along the centerline of the annular experimental resonator [$r=R_0$, s, $z=0$, see Fig. 1(a)] of the acoustic particle velocity is presented in Figs. 2(a) and 2(b) for a phase shift ϕ between the membrane displacements of the loudspeakers of 0 and $-\pi/2$, respectively. More precisely, the normalized distributions of the axial velocity fields $|v_s(R_0,s,0)/\langle v_s(R_0,s,0)\rangle_s|$, where $\langle v_s(R_0,s,0)\rangle_s$ denotes the average amplitude of acoustic velocity along the s-coordinate, are presented: full lines show the predicted spatial distribution [Eqs. (7) and (8)] and squares show the measurement results. The experimental and theoretical results are in good agreement. When $\phi = 0$ [Fig. 2(a)], the spatial distribution of the acoustic field in the resonator is the one of a standing wave. When $\phi=-\pi/2$ [Fig. 2(b)], the results show

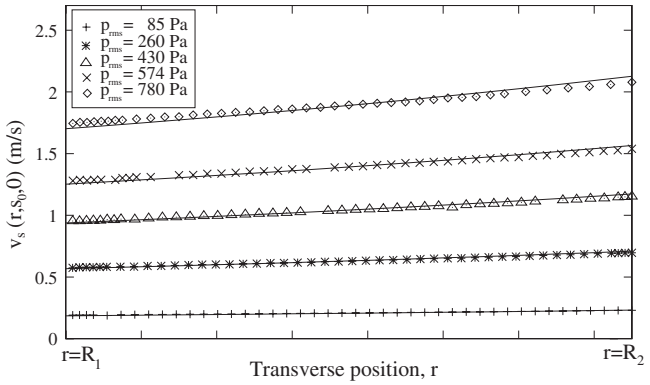


FIG. 3. Spatial distribution of the amplitude of the axial acoustic velocity $v_s(r,s_0,0)$, measured by LDV at position $s_0=0.72\ \mathrm{m}\approx 3L/8$ for different amplitudes of acoustic pressure, with respect to the r-coordinate. Continuous lines correspond to the calculated distribution [Eq. (12)].

that the measured amplitude of the acoustic particle velocity (squares: ⊡) is nearly constant along the s-coordinate, as predicted by the theory (full line).

In Fig. 3, the measured spatial distribution of acoustic velocity with respect to coordinate r is shown for different amplitudes of acoustic pressure. The phase shift between the two drivers is set to $\phi=-\pi/2$. The LDV measurements are performed at position $s_0=0.72\ \mathrm{m}\approx 3L/8$. It was technically impossible here to measure this acoustic particle velocity inside the thermo-viscous boundary layers, so that the spatial distribution in Fig. 3 does not show that the particle velocity vanishes near the walls of the resonator, at positions $r=R_1$ and $r=R_2$. It appears clearly that the particle velocity is not uniform across a section of the waveguide. This effect is due to the curvature of the resonator, and an estimate of the actual velocity distribution can be obtained by assuming that, in the closed-loop resonator of Fig. 1(a), the pressure wave is a plane wave traveling in the $(+s)$ direction without dissipation $[\tilde{p}(r,s,z)=\tilde{p}(s)=\tilde{p}(s_0)e^{ik_0(s-s_0)}]$ so that, using Euler's equation in the cylindrical system of coordinates (r,s,z), the complex amplitude of the axial acoustic particle velocity

$$\tilde{v}_s(r,s_0,z)=\frac{1}{i\rho_0\omega r}\frac{\partial\tilde{p}(s_0)}{\partial s} \tag{12}$$

is proportional to $(1/r)\partial_s\tilde{p}$. The corresponding transverse distribution of the acoustic particle velocity is also presented in Fig. 3 (continuous lines) and it is in good agreement with the results of measurements.

C. Outer acoustic streaming velocity measurements

Acoustic streaming velocity measurement by LDV is a tricky procedure in which several parameters such as the number of seeding particles getting through the measurement volume and their influence on the mean density of fluid, the temperature, and the static pressure variations impact the experimental results. The most annoying difficulty that we have encountered is the necessity to wait for quite a long time after the introduction of seeding particles (wood smoke) before to proceed to reliable acoustic streaming velocity measurements. Indeed, while the measured acoustic particle velocity converges to a constant value only several seconds after the introduction of seeding particles, the streaming velocity reaches its steady-state value after approximately 30 min, the initial streaming velocity value being one order of magnitude higher than its steady-state value. It should be mentioned that such an effect has already been reported by Thompson *et al.*[23] and confirmed by Moreau *et al.*[24] The most plausible reason which can explain this effect is the one invoked by Thompson *et al.*, who have clearly demonstrated that the thermal boundary condition imposed on the walls of the resonator has a strong influence on both the characteristic time necessary to reach the steady-state acoustic streaming velocity and on the amplitude of the steady-state acoustic streaming velocity. In the present experiments, no thermal conditions are imposed on the resonator (this is referred to as "uncontrolled" boundary condition by Thompson *et al.*). The acoustic streaming velocity reaches its steady-state value within 27 min (the time required for the acoustic streaming velocity to reach 95% of its steady-state value), which is twice longer than the time of 14 min reported by Thompson *et al.* In the present device, this long time delay probably corresponds to the time for stabilization of a heterogeneous temperature distribution, which is due to various heat sources (notably the loudspeakers). Actually, this thermal equilibrium time is controlled by the convective heat transport due to the acoustic streaming which itself is controlled by the temperature distribution. In practice, this means that after each introduction of seeding particles into the waveguide, LDV measurements cannot begin before this 27 min time delay, and have to be stopped after approximately 1 h due to the gradual decrease in the number of seeding particles getting through the measurement volume, with subsequent decrease in the signal to noise ratio.

The measurement of the spatial distribution of the outer ("outer" meaning far from the boundary layers) acoustic streaming velocity $v_{sm}(R_0,s,0)$ along the centerline is presented in Fig. 4 for both $\phi=0$ and $\phi=-\pi/2$. It is important to note that it was not possible to prevent from annoying effects due to the loudspeakers in this device. This was not the case in the devices studied by Thompson *et al.*[22] and Moreau *et al.*:[24] Thompson *et al.* took care to proceed to the measurements far from the acoustic sources (approximately 1 m) while Moreau *et al.* took care to design their waveguide in such a way that separation effects due to the geometrical singularities in the vicinity of the loudspeakers are minimized. In the present device, the disturbances due to the loudspeakers impact the measurement of acoustic streaming along almost half the length of the resonator (approximately from $s\approx 5L/8$ to $s\approx L/8$). However, far away from the sources, after a stabilization distance of about 40 cm, we consider that the observed streaming is the acoustic streaming itself (i.e., the streaming resulting from nonlinear acoustic effects in the vicinity of the resonator walls). In Fig. 4, both the acoustic particle velocity magnitude [Fig. 4(a)] and the acoustic streaming velocity magnitude [Fig. 4(b)] are represented along the waveguide. No matter how the phase shift between the loudspeakers is adjusted ($\phi=0$ or $\phi=-\pi/2$), the spatial zone where the measurements of acoustic streaming may be considered as reliable extends from $s\approx L/8$ to $s\approx 5L/8$. In this region, and in the case when a

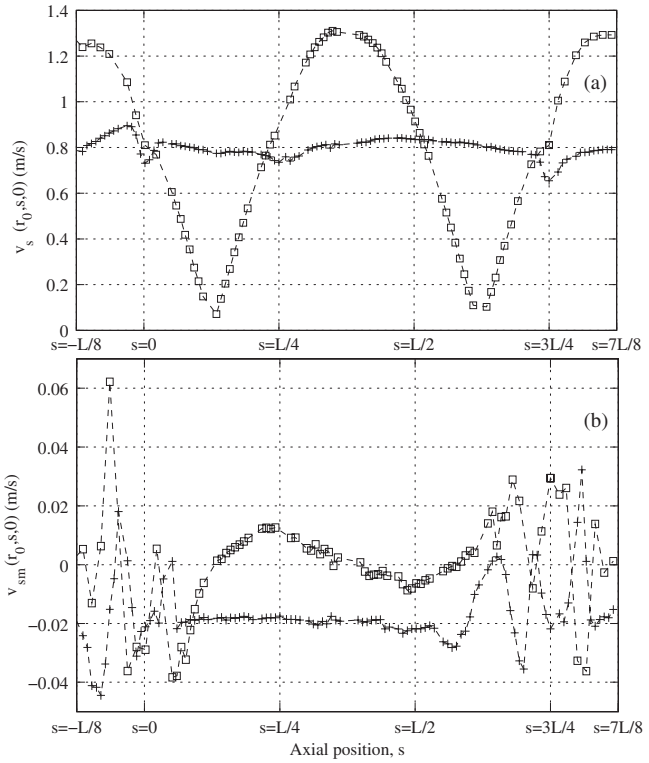


FIG. 4. (a) Spatial distribution of the amplitude of the axial acoustic velocity $v_s(R_0,s,0)$ and (b) of the outer acoustic streaming velocity $v_{sm}(R_0,s,0)$ with respect to the axial position s, measured by LDV for a phase shift $\phi=-\pi/2$ (crosses: ×) and $\phi=0$ (squares: ⊡). In both cases ($\phi=0$ and $\phi=-\pi/2$), the root-mean-square amplitude of the electric voltage applied to the loudspeakers is 1.8 V.

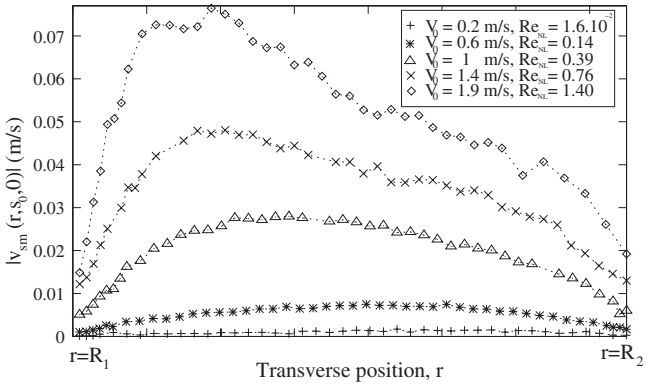


FIG. 5. Spatial distribution of the absolute values of the amplitude of the axial outer acoustic streaming velocity $|v_{sm}(r,s_0,0)|$, measured by LDV at position $s_0=0.72\ \text{m}\approx 3L/8$ for a phase shift $\phi=-\pi/2$, with respect to the r-coordinate and for different values of the typical acoustic particle velocity $V_0=\langle v_s(r,s_0,0)\rangle_r$.

standing wave is excited by the loudspeakers ($\phi=0$, squares: ⊡), it is interesting to note from the simultaneous observations of acoustic and acoustic streaming velocities that our results match the classical observations of acoustic streaming in standing wave resonators. There is indeed a periodic variation in acoustic streaming velocity, which cancels at the locations of acoustic particle velocity nodes and antinodes. Moreover, it is also noticeable that the acoustic streaming flow along the centerline of the waveguide is directed toward acoustic velocity antinodes. In the case when $\phi=-\pi/2$ (traveling wave, crosses: ×), the amplitude of the axial streaming velocity $v_{sm}(R_0,s,0)$ is roughly constant and negative. This observed negative sign is in accordance with the theoretical results (see Sec. III B) which predict that, in the present case for which $\delta_\nu/a\approx 2\times 10^{-3}\ll 1$, acoustic streaming is directed opposite to the traveling wave.

The spatial distribution of the outer acoustic streaming velocity with respect to the r-coordinate is presented in Fig. 5. Measurements are performed at position $s_0=0.72$ m $\approx 3L/8$ and for different amplitudes of acoustic particle velocity. It clearly appears that the transverse distribution of the acoustic streaming velocity is not symmetrical with respect to the centerline. It is also noticeable that the maximum of streaming velocity is shifted toward the external wall [located at $r=R_1$, see Fig. 1(a)] of the resonator as the magnitude of the acoustic particle velocity increases. This behavior might be attributed to the effect of fluid inertia which leads to acoustic streaming distortion.[14] In order to evaluate if the influence of fluid inertia on acoustic streaming is important or not, a nonlinear a dimensional Reynolds number is defined as[14]

$$\text{Re}_{\text{NL}}=\left(\frac{V_0}{c_0}\right)^2\left(\frac{a}{2\delta_\nu}\right)^2, \tag{13}$$

where V_0 is a typical value of acoustic velocity at position $s_0=0.72\ \text{m}\approx 3L/8$ (in the following, $V_0=\langle v_s(r,s_0,0)\rangle_r$ and $\langle\cdots\rangle_r$ denotes taking the spatial average over the r-coordinate). This number Re_{NL}, which was first introduced by Menguy and Gilbert,[14] is an appropriate dimensionless number which states the limit between slow and fast streaming in the kind of device considered here (i.e., frequencies in the audible range, radii of order a centimeter, acoustic levels up to 160 dB). If $\text{Re}_{\text{NL}}\ll 1$, the corresponding streaming is called slow streaming and the effect of inertia is assumed to be negligible (so that the model presented in Appendix is valid). If $\text{Re}_{\text{NL}}\geq 1$, the streaming is called fast streaming (or nonlinear streaming) and fluid inertia may influence the generation of acoustic streaming, notably by distorting the shape of the transverse distribution of the acoustic streaming velocity. In the experimental results presented in Fig. 5, the parameter Re_{NL} varies from 1.6×10^{-2} to 1.4 so the effect of fluid inertia may be responsible for the observed asymmetry of the acoustic streaming transverse distribution. When the average amplitude of acoustic particle velocity V_0 exceeds 1 m/s (i.e., when $\text{Re}_{\text{NL}}\geq 0.4$), the transverse distribution of acoustic streaming indeed becomes asymmetrical. Another plausible reason which may explain these results is that the curvature of the resonator, which impacts the transverse profile of acoustic velocity, also impacts the transverse profile of acoustic streaming velocity.

The average value $\langle v_{sm}(r,s_0,0)\rangle_r$ over the r-coordinate of the acoustic streaming velocity performed at position $s_0=0.72\ \text{m}\approx 3L/8$ is presented in Fig. 6 as functions of p_{rms}^2, where p_{rms} is the root-mean-square of the acoustic pressure amplitude measured at position $s=0.75$ m. The continuous line corresponds to the linear fit of the experimental data ($\langle v_{sm}\rangle_r=\alpha_{\text{fit}}p_{\text{rms}}^2$). It is verified that the acoustic streaming velocity is proportional to acoustic intensity, and the calcu-

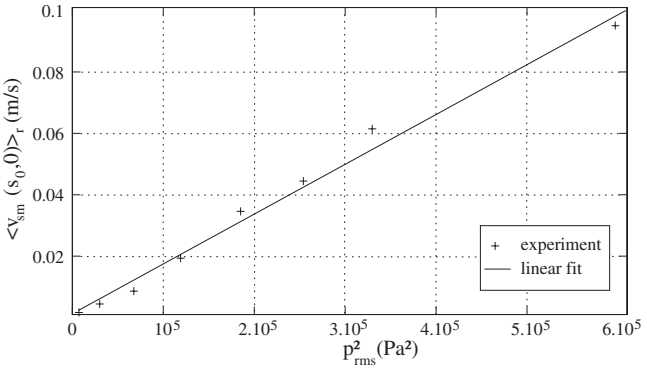


FIG. 6. Average values $\langle v_{sm}(r,s_0,0)\rangle_r$ of the amplitude of the axial outer acoustic streaming velocity $|v_{sm}(r,s_0,0)|$ over the r-coordinate, at position $s_0=0.72\ \mathrm{m}\approx 3L/8$, with respect to the square of the root-mean-square amplitude of the acoustic pressure p_{rms}^2.

lated slope of the linear fit is $\alpha_{\mathrm{fit}}\approx -1.6\times 10^{-7}\ \mathrm{m\ s^{-1}\ Pa^{-2}}$. This value is very close to the predicted value $\alpha\approx -1.8\times 10^{-7}\ \mathrm{m\ s^{-1}\ Pa^{-2}}$ obtained from our simplified theory [Eq. (11)]. It is worth noting that, though the results obtained exhibit a complicated structure of acoustic streaming with a noticeable effect of fluid inertia and maybe of the curvature of the resonator, the simplified analytical predictions are not so far from experiments.

V. CONCLUSION

In this paper, an experimental study of a traveling wave closed-loop resonator is presented and analytical interpretations are suggested. To our knowledge, this is the first experimental study which focuses on this original device and specifically on the development of acoustic streaming generated by a resonant traveling acoustic wave. The distribution of the acoustic field, which is controlled both by the relative driving amplitude and phase shift between two drivers, is measured using LDV and the results appear to be coherent with the analytical results. The experimental analysis of the acoustic streaming development is also carried out using LDV. The obtained results show that the effects of fluid inertia and maybe the effect of the curvature of the resonator impact the spatial distribution of acoustic streaming through the cross-section of the resonator. It is also demonstrated that the measured cross-sectional average streaming velocity is in good agreement with the value predicted by our simplified theoretical model. It would be interesting in the future to repeat these experiments in the presence of a controlled temperature gradient and to compare the obtained results with analytical predictions.

ACKNOWLEDGMENTS

This work has been supported by the French National Research Agency (Contract No. ANR-05-BLANC-0016-2). The authors would like to thank Solenn Moreau, Hélène Bailliet, David Marx, and Jean-Christophe Valière for their helpful assistance and numerous advices concerning LDV measurements. The authors are also indebted to Michel Bruneau and Vitalyi Gusev for helpful discussions.

APPENDIX: ESTIMATE OF THE ACOUSTIC STREAMING VELOCITY

An order of magnitude estimate of the acoustic streaming velocity can be obtained by using the model described in Ref. 18. For the sake of simplicity, it is assumed here that the unwrapped resonator has a circular cross-section of hydraulic radius $a/2$ (calculations would be more complicated to carry on in the case of a square cross-section). The coordinate system necessary to describe the fluid motion thus consists of an axial coordinate y running along the centerline of the unwrapped resonator and a radial coordinate r running perpendicular to the centerline (with $r=0$ along the centerline). The fluid motion is assumed to be symmetric about the centerline. Assuming that the boundary layer approximation is valid, the equation describing the transverse variations of the axial streaming velocity v_{ym} is obtained from a successive approximations approach,[13,18] leading to

$$\nu_0\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial v_{ym}}{\partial r}\right)=\frac{1}{\rho_0}\frac{\partial}{\partial y}(p_h+\rho_0\overline{v_y^2})+\frac{1}{r}\frac{\partial}{\partial r}(r\overline{v_y v_r})-\beta\frac{\nu_0}{T_0}\frac{1}{r}\frac{\partial}{\partial r}\left(r\overline{\tau\frac{\partial v_y}{\partial r}}\right),\tag{A1}$$

where $\overline{\cdots}$ is used to denote time averaging, τ denotes acoustic fluctuations of temperature, ν_0 is the fluid viscosity evaluated at temperature T_0 ($\nu_0\propto T_0^{\beta}$, with $\beta=0.73$),[13] p_h is the hydrodynamic pressure accompanying the streaming, and where $v_y(y,r,t)$ and $v_r(y,r,t)$ are the axial and transverse acoustic velocities, respectively. Introducing the dimensionless coordinate $\eta=(2r)/a$, and notifying that the cross-sectional average mass flow $\langle M\rangle$ is necessarily constant ($\langle\cdots\rangle=2\int_0^1\ldots\eta d\eta$) due to the closed-loop geometry, it is possible to eliminate the hydrodynamic pressure p_h in Eq. (A1) and to obtain the following expression of the acoustic streaming velocity:

$$\langle v_{ym}(y)\rangle=\frac{1}{\rho_0}(\langle M\rangle-\langle\overline{\rho v_y}\rangle),\tag{A2}$$

where ρ denotes acoustic fluctuations of fluid density. The cross-sectional average mass flow $\langle M\rangle$ is given by

$$\langle M\rangle=\left(\oint m(y)dy\right)\Big/(4L\nu_0/a^2),\tag{A3}$$

where the function $m(y)$, which represents a density of sources inducing acoustic streaming, is given by

$$m(y)=\left\langle\frac{\rho_0}{a}\int_1^{\eta}\left[\Re(\tilde{v}_y\tilde{v}_\eta^*)-2\beta\frac{\nu_0}{T_0}\frac{1}{a}\Re\left(\tilde{\tau}\frac{\partial\tilde{v}_y^*}{\partial\eta}\right)\right]\cdot d\eta'+2\frac{\nu_0}{a^2}\Re(\tilde{\rho}\tilde{v}_y^*)\right\rangle,\tag{A4}$$

where the relation $\overline{gh}=(1/2)\Re(\tilde{g}\tilde{h}^*)$ (* denoting complex conjugate) is used to calculate the time average of the product gh, $\tilde{g}$ and $\tilde{h}$ being the complex amplitudes of the arbitrary functions g and h.

Moreover, the acoustic variables $\tilde{\tau}$, $\tilde{v}_y$, $\tilde{v}_\eta$, and $\tilde{\rho}$ can be expressed as a function of the acoustic pressure $\tilde{p}$ as follows:[18]

$$\tilde{v}_y = \frac{1 - F_\nu}{i\omega\rho_0}\frac{d\tilde{p}}{dy}, \tag{A5}$$

$$\tilde{\tau} = \frac{1 - F_\kappa}{\rho_0 C_p}\tilde{p}, \tag{A6}$$

$$\tilde{\rho} = \frac{1 + (\gamma - 1)F_\kappa}{c_0^2}\tilde{p}, \tag{A7}$$

$$\tilde{v}_\eta = -\frac{a}{4i\omega\rho_0}\left(\frac{d}{dy}\left[(\eta - \Phi_\nu)\frac{d\tilde{p}}{dy}\right] + \left(\frac{\omega}{c_0}\right)^2(\eta + (\gamma - 1)\Phi_\kappa)\tilde{p}\right), \tag{A8}$$

where C_p is the isobaric specific heat of fluid. In the present case of a circular cross-section, the functions $F_{\nu,\kappa}$ and $\Phi_{\nu,\kappa}$ are given by[19]

$$F_{\kappa,\nu} = \frac{J_0(b_{\kappa,\nu}\eta)}{J_0(b_{\kappa,\nu})}, \tag{A9}$$

$$\Phi_{\kappa,\nu} = \frac{2}{b_{\kappa,\nu}}\frac{J_1(b_{\kappa,\nu}\eta)}{J_0(b_{\kappa,\nu})}, \tag{A10}$$

where $b_{\kappa,\nu}=((1+i)a)/(2\delta_{\kappa,\nu})$, and where J_k are the cylindrical Bessel functions of the first kind and order k.

As mentioned in Sec. III, it is possible to calculate the spatial distribution of acoustic pressure $\tilde{p}(y)$ in the entire resonator, which can be reported in Eqs. (A5)–(A8) in order to calculate the average cross-sectional mass flow $\langle M\rangle$. Using Eq. (A2), the cross-sectional streaming velocity $\langle v_{ym}\rangle$ is finally obtained. This acoustic streaming velocity depends actually on the axial coordinate y, but in the case of a quasi-traveling wave (when the displacements of the two drivers are $\pi/2$ out of phase), it is almost constant. Indeed, after some calculations using the dimensions of the experimental device (a=7.5 cm, L=2.12 m), it is found that the average streaming velocity $1/L\oint\langle v_{ym}\rangle\cdot dy$ along the closed-loop waveguide is proportional to the square of the modulus of acoustic pressure amplitude as follows:

$$\frac{1}{L}\oint\langle v_{ym}\rangle\cdot dy \approx -\Gamma\left|\frac{1}{L}\oint\tilde{p}(y)\cdot dy\right|, \tag{A11}$$

with $\Gamma \approx 3.6\times10^{-7}$ m s^{-1} Pa^{-2} and that the maximum variation of the cross-sectional average streaming velocity $\langle v_{ym}\rangle$ along the resonator is lower than 0.5% of $1/L\oint\langle v_{ym}\rangle\cdot dy$.

[1] J. W. Strutt (Lord Rayleigh), "On the circulation of air observed in Kundt's tube, and on some allied acoustical problems," Philos. Trans. R. Soc. London, Ser. A **36**, 10–11 (1883).
[2] P. Vainshtein, M. Fishman, and C. Gutfinger, "Acoustic enhancement of heat transfer between two parallel plates," Int. J. Heat Mass Transfer **38**, 1893–1899 (1995).
[3] R. Moroney and R. White, "Microtransport induced by ultrasonic lamb waves," Appl. Phys. Lett. **59**, 774–776 (1991).
[4] N. Nguyen and R. White, "Design and optimization of an ultrasonic flexural plate wave micropump using numerical simulations," Sens. Actuators, A **77**, 229–236 (1999).
[5] P. Luchini and F. Charru, "Acoustic streaming past a vibrating wall," Phys. Fluids **17**, 122106 (2005).
[6] J. Hu, K. Nakamura, and S. Ueha, "An analysis of a noncontact ultrasonic motor with ultrasonically levitated motor," Ultrasonics **35**, 459–467 (1997).
[7] G. W. Swift, *Thermoacoustics: A Unifying Perspective for Some Engines and Refrigerators* (Acoustical Society of America, Melville, NY, 2002).
[8] G. Penelet, V. Gusev, P. Lotton, and M. Bruneau, "Experimental and theoretical study of processes leading to steady-state sound in annular thermoacoustic prime movers," Phys. Rev. E **72**, 016625 (2005).
[9] G. Penelet, V. Gusev, P. Lotton, and M. Bruneau, "Non trivial influence of acoustic streaming on the efficiency of annular thermoacoustic prime movers," Phys. Lett. A **351**, 268–273 (2006).
[10] G. W. Swift, D. Gardner, and S. Backhaus, "Acoustic recovery of lost power in pulse tube refrigerators," J. Acoust. Soc. Am. **105**, 711–724 (1999).
[11] G. W. Swift and S. Backhaus, "A thermoacoustic Stirling heat engine," Nature (London) **399**, 335–338 (1999).
[12] J. R. Olson and G. W. Swift, "Acoustic streaming in pulse tube refrigerators: Tapered pulse tube," Cryogenics **37**, 769–776 (1997).
[13] N. Rott, "The influence of heat conduction on acoustic streaming," ZAMP **25**, 417–421 (1974).
[14] L. Menguy and J. Gilbert, "Non-linear acoustic streaming accompanying a plane stationary wave in a guide," Acust. Acta Acust. **86**, 249–259 (2000).
[15] M. F. Hamilton, Y. Ilinskii, and E. Zabolotskaya, "Thermal effects of acoustic streaming in standing waves," J. Acoust. Soc. Am. **114**, 3092–3101 (2003).
[16] M. Amari, V. Gusev, and N. Joly, "Temporal dynamics of the sound wind in acoustitron," Acust. Acta Acust. **89**, 1008–1024 (2003).
[17] D. Gedeon, "DC gas flows in Stirling and pulse-tube cryocoolers," Cryocoolers **9**, 385–392 (1997).
[18] V. Gusev, S. Job, H. Bailliet, P. Lotton, and M. Bruneau, "Acoustic streaming in annular thermoacoustic prime-movers," J. Acoust. Soc. Am. **108**, 934–945 (2000).
[19] H. Bailliet, V. Gusev, R. Raspet, and R. Hiller, "Acoustic streaming in closed thermoacoustic devices," J. Acoust. Soc. Am. **110**, 1808–1821 (2001).
[20] M. Mironov, V. Gusev, Y. Auregan, P. Lotton, and M. Bruneau, "Acoustic streaming related to minor loss phenomenon in differentially heated elements of thermoacoustic devices," J. Acoust. Soc. Am. **112**, 441–445 (2002).
[21] T. Biwa, Y. Tashiro, M. Ishigaki, Y. Ueda, and T. Yazaki, "Measurements of acoustic streaming in a looped-tube thermoacoustic engine with a jet pump," J. Appl. Phys. **101**, 64914 (2007).
[22] M. Thompson and A. Atchley, "Simultaneous measurement of acoustic and acoustic streaming velocities in a standing wave using laser Doppler Anemometry," J. Acoust. Soc. Am. **117**, 1828–1838 (2005).
[23] M. Thompson, A. Atchley, and M. Maccarone, "Influences of a temperature gradient and fluid inertia on acoustic streaming in a standing wave," J. Acoust. Soc. Am. **117**, 1839–1849 (2005).
[24] S. Moreau, H. Bailliet, and J. C. Valière, "Measurements of inner and outer streaming vortices in a standing waveguide using laser Doppler velocimetry," J. Acoust. Soc. Am. **123**, 640–647 (2008).
[25] S. Backhaus, E. Tward and M. Petach, "Traveling-wave thermoacoustic electric generator," Appl. Phys. Lett. **85**, 1085–1087 (2004).
[26] M. Amari, V. Gusev, and N. Joly, "Transient unidirectional acoustic streaming in annular resonators," Ultrasonics **42**, 573–578 (2004).
[27] P. Ceperley, "Split mode travelling wave ring-resonator," U.S. Patent No. 4,686,407 (1987).
[28] C. C. Lawrenson, L. D. Lafleur, and F. D. Shields, "The solution for the propagation of sound in a toroidal waveguide with driven walls (the acoustitron)," J. Acoust. Soc. Am. **103**, 1253–1260 (1998).
[29] P. Ceperley, "Rotating waves," Am. J. Phys. **60**, 938–942 (1992).
[30] W. P. Arnott, H. E. Bass, and R. Raspet, "General formulation of thermoacoustics for stacks having arbitrarily shaped pore cross sections," J. Acoust. Soc. Am. **90**, 3228–3237 (1991). Equation (6) is obtained by using the expression (73) in the above mentioned reference, and by assuming $\delta_{\nu,\kappa} \ll a$.
[31] L. Simon, O. Richoux, A. Degroot, and L. Lionet, "Laser Doppler velocimetry for joint measurements of acoustic and mean flow velocities: LMS-based algorithm and CRB calculation," IEEE Trans. Instrum. Meas. **57**, 1455–1464 (2008).

Acoustic measurement of bubble size in an inkjet printhead

Roger Jeurissen and Arjan van der Bos
Physics of Fluids Group, Faculty of Science and Technology and Burgers Center of Fluid Dynamics, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

Hans Reinten, Marc van den Berg, Herman Wijshoff, and Jos de Jong
Océ Technologies B.V., P.O. Box 101, 5900 MA Venlo, The Netherlands

Michel Versluis and Detlef Lohse
Physics of Fluids Group, Faculty of Science and Technology and Burgers Center of Fluid Dynamics, MESA+ Institute for Nanotechnology and Institute of Mechanics, Processes and Control-Twente (IMPACT), University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

(Received 30 April 2009; revised 14 August 2009; accepted 17 August 2009)

The volume of a bubble in a piezoinkjet printhead is measured *acoustically*. The method is based on a numerical model of the investigated system. The piezo not only drives the system but it is also used as a sensor by measuring the current it generates. The numerical model is used to predict this current for a given bubble volume. The inverse problem is to infer the bubble volume from an experimentally obtained piezocurrent. By solving this inverse problem, the size and position of the bubble can thus be measured acoustically. The method is experimentally validated with an inkjet printhead that is augmented with a glass connection channel, through which the bubble was observed optically, while at the same time the piezocurrent was measured. The results from the acoustical measurement method correspond closely to the results from the optical measurement.
 [DOI: 10.1121/1.3224760]

PACS number(s): 43.25.Yw, 43.38.Fx, 43.35.Yb, 43.30.Zk [CCC] Pages: 2184–2190

I. INTRODUCTION

The dynamics of a sound driven free bubble in infinite volume is well described by the Rayleigh–Plesset equation,[1–3] whose validity even under the extreme conditions of single bubble sonoluminescence has been thoroughly established.[4] However, many important cases of bubble dynamics occur under constraint conditions, in finite volumes of liquid, rather than infinite volumes, such as in confined spaces and near a wall.[5–8] Examples include the behavior of gas bubbles in blood vessels, aiming at improving ultrasound diagnostics and treatment,[9] or thermal inkjet printing and other microfluidic applications, where bubbles are used as actuators.[10,11] However, bubbles can also disrupt the operation of the printhead as was shown in earlier research.[12–14] Although inkjet printing is a robust process and billions of droplets can be printed without problems, there is a small chance that during actuation a small air bubble is entrapped at the nozzle of an ink channel. The bubble influences the channel acoustics, reducing the pressure buildup at the nozzle. The bubble grows by rectified diffusion until it reaches a diffusive equilibrium.[12–14] At this size, the pressure buildup at the nozzle is insufficient for droplet production, so that the nozzle fails. This malfunctioning can be detected acoustically,[13] but until now the relation between bubble size and channel acoustics has not been shown quantitatively. In fact, in many studies, the bubble was assumed to behave as if it were in an unbounded liquid.[15,16]

The dynamics of a bubble in confined space is fundamentally different from that in an infinite volume of liquid where the far field is three dimensional. In contrast, in a compressible inviscid liquid, the far field of a bubble between two parallel infinite walls is two dimensional,[17] and the far field of a bubble in an infinitely long pipe is one dimensional.[18–20] An incompressible liquid does not allow bubble volume fluctuations in either confined space, while the volume fluctuations in an unbounded volume of liquid are possible and governed by the Rayleigh–Plesset equation. Models that assume an unbounded volume of liquid are therefore inappropriate for a bubble in a confined space.

In this study, a model is used that captures the effect that a bubble has on the channel acoustics and vice versa. To validate the model, experimental results are presented which correlate the acoustic change inside the channel with optical measurements of an entrained air bubble.

II. GEOMETRY OF THE INKJET PRINTHEAD

The inkjet printhead that is used in this research is developed by Océ Technologies B.V. This experimental printhead consists of 256 similar ink channels. Each channel has an actuator section with a rectangular cross section of 118 $\times 218$ μm^2 and a length of 8 mm. A cylindrical channel section with a radius of 125 μm and a length of 1.5 mm connects the actuator section to the nozzle. A piezo is placed onto the actuator section. When a trapezoidal pulse of 13 μs (4 μs rise time, 5 μs plateau, and 4 μs fall time)[21] is applied to this piezo, it generates acoustic waves in the channel. The generated waves travel through the channel and are reflected at the ink reservoir at one side, and at the nozzle at the other side. The result is a velocity and pressure buildup at the nozzle which leads to a droplet being ejected.[22,23] Typically, droplets of 30 pl are generated at a rate of 20 kHz with a velocity of 6 m/s.

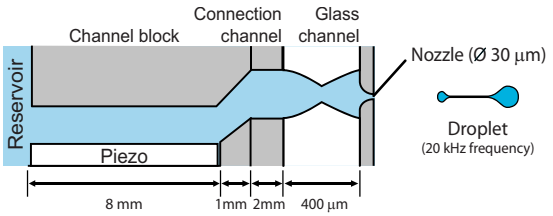


FIG. 1. (Color online) The channel inside the printhead is about 10 mm long and is actuated by an 8 mm long piezo. In between the channel block and the nozzle plate, a 400 μm long glass connection channel is placed through which the bubble dynamics can be observed.

To visualize the dynamics of the entrained air bubble, a 400 μm long glass connection channel (Micronit Microfluidics B.V., The Netherlands, info@micronit.com) was interposed between the ink channel and the nozzle plate (Fig. 1), similarly as done in Ref. 13. This channel was made by powder blasting which resulted in an hourglass shape with a waist diameter of 220 μm and a maximum diameter of 300 μm at the ends. To interpose the glass connection channel an extra 2 mm connection channel was required. In Fig. 2, the connection channel with an air bubble inside is shown. On top of the connection channel, a 100 μm thick nickel nozzle plate is glued. The trumpet shaped nozzles have a diameter of 30 μm at the exit and 130 μm at the inlet.

III. EXPERIMENTAL PARAMETERS

Besides visualizing the bubble dynamics, also the pressure variations inside the channel were measured. This was done by measuring the piezocurrent. This technique[24] has earlier been applied in Ref. 13. Even small pressure fluctuations in the channel result in measurable current fluctuations from the piezo. As this signal is only measured in between the actuation pulses, the time window where the current can be measured is 37 μs at a droplet production rate of 20 kHz. An example of this piezocurrent is shown in Fig. 3. This figure illustrates that the acoustic signal changes significantly when the channel acoustics are disturbed by air entrapment.

The piezocurrent was measured at a range of bubble volumes. To accomplish this, air entrapment was induced by physically blocking a channel while actuating. The actuation was continued until the entrapped bubble reached its diffusive equilibrium size, which is about 120 pl. Then, the actuation was stopped allowing the bubble to dissolve. The bubble dissolves at a rate of approximately 0.5 pl/s, so it takes about 4 min for a 120 pl bubble to fully dissolve. During the dissolution of the bubble, piezocurrent data were gathered by actuating at a frequency of 1 Hz. At this reduced actuation rate, rectified diffusion is not strong enough to sustain the bubble, so it dissolves. 1 μs before every actuation pulse, an image of the bubble was captured. In this way, motion blur due to volume oscillations was prevented.

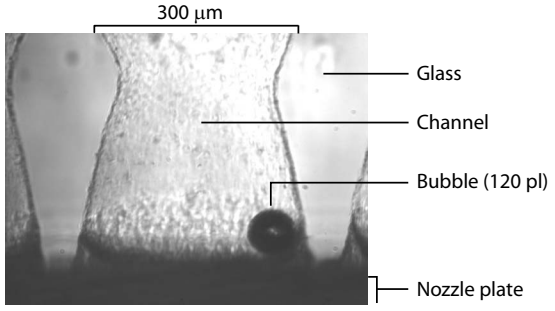


FIG. 2. A microscope image showing an entrapped air bubble in the glass connection channel. While actuating, the fully grown air bubble will just remain oscillating in the channel indefinitely. Note the position of the air bubble: due to the secondary Bjerknes force, it is pushed against the glass wall where it stays fixed even after the actuation is stopped. On the left and right sides of the channel, the neighboring channels can also be seen.

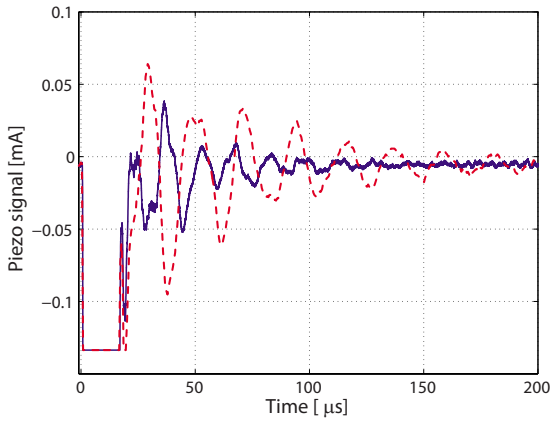


FIG. 3. (Color online) Piezocurrent of a normal operating nozzle (solid) and with an entrained air bubble with a volume of V_b=80 pl (dashed) close to the nozzle plate. It can be seen in this figure that the volume oscillations of the entrapped bubble modify the piezocurrent significantly; the piezocurrent amplitude is less damped and the main frequency decreases.

IV. MODELING THE PRINTHEAD

Deformation of a piezo gives rise to a current I from the actuator. Such a deformation can be caused by varying the voltage over the electrodes. Thanks to this effect, the piezo can be used as an actuator. Another way in which the piezo can be deformed is caused by acoustic waves in the channel. Therefore the piezoelement can be used also as a sensor. The piezocurrent is calculated by using the model developed in Ref. 14, which links the Rayleigh–Plesset equation to the equations that govern the propagation of acoustic waves in a viscous medium in a flexible pipe and the response of the piezo and channel to the actuator voltage.

Acoustically, the printhead consists of four linked sections of pipes, as shown in Fig. 4. The properties of the channel are constant over each section. The relevant properties are the piezoelectric expansion coefficient α_j, the wall flexibility β_j, the cross sectional area A_j, the velocity of sound in the liquid c, the liquid density ρ, the viscosity μ, and the length L_j of the channel section. The piezoelectric expansion coefficient is defined as

$$\alpha \equiv \frac{1}{A}\left(\frac{\partial A}{\partial U}\right)_P, \tag{1}$$

where U is the voltage over the electrodes of the piezoelement and P is the pressure in the channel. The wall flexibility is defined as

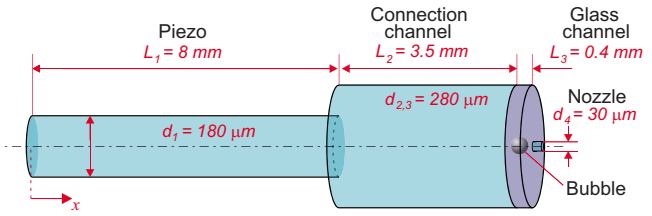


FIG. 4. (Color online) The printhead as it is implemented in the model. From left to right: actuator channel, connection channel, glass connection channel, and nozzle.

$$\beta \equiv \frac{1}{A}\left(\frac{\partial A}{\partial P}\right)_U. \tag{2}$$

These quantities can be determined with a solid mechanics calculation, provided that the geometry and material parameters are accurately known. They can also be determined by measuring the piezocurrent in the absence of a bubble.

The analysis is performed in the frequency domain. The discrete Fourier transform is defined through

$$f(t) = \sum_j F(\omega_j) e^{i\omega_j t}, \tag{3}$$

where $f(t)$ is the relevant quantity in the time domain, and $F(\omega)$ is the corresponding quantity in the frequency domain. The explicit dependence on frequency is dropped in the remainder of the paper for the sake of brevity. The pressure P is decomposed into the waves propagating to the left P_l and right P_r. For each channel section, the amplitudes of the left and right propagating waves are calculated per frequency,

$$P = \sum_j P_r e^{i(\omega_j t - kx)} + P_l e^{i(\omega_j t + kx)} + P_s. \tag{4}$$

The pressure P_s due to the actuator depends only on the imposed actuator voltage. The wave number k is a complex quantity due to viscous dissipation. For a cylindrical pipe, a closed form expression can be obtained analytically,[25] namely,

$$k = \frac{\omega}{c_{\text{eff}}}\sqrt{\frac{1}{1 + \frac{2\sqrt{-i}J_1(\text{Woi}^{3/2})}{\text{Wo}J_0(\text{Woi}^{3/2})}}}. \tag{5}$$

The functions J_0 and J_1 are the ordinary Bessel functions of the first kind, of zeroth and first orders, respectively. The Womersley number Wo is the ratio of the inertia of the oscillating velocity field over the viscosity,

$$\text{Wo}_j = \frac{1}{2} d_j \sqrt{\frac{\omega\rho}{\mu}}, \tag{6}$$

where d_j is the diameter of the section. The effective wave velocity is the inviscid phase velocity of acoustic waves. This quantity differs from the velocity of sound due to wall flexibility. It was derived by Young[26] and is given by

$$c_{\text{eff}} = \sqrt{\frac{c^2}{1 + \beta c^2 \rho}}. \tag{7}$$

If the wall flexibility β vanishes, the effective wave velocity is equal to the velocity of sound c. If the walls are flexible the effective wave velocity is smaller. The wave number (5) has been the main result of the acoustical model.[25] The boundary conditions are continuity of pressure and volume flow rate. Equation (4) for the pressure, Eq. (5) for the wave number, and the boundary conditions describe the propagation of acoustic waves in a flexible channel filled with a viscous liquid. The effect of the bubble is included in the volume flow rate balance between the nozzle and the glass connection channel by explicitly considering the flux q_b from the bubble,

$$A_c u_c = A_n u_n + q_b. \tag{8}$$

Here A_c and A_n are the cross sectional areas of the glass connection channel and the area of the nozzle, respectively, and u_c and u_n are fluid velocities in these sections. The volume flux from the bubble is calculated with the Rayleigh–Plesset equation

$$r\ddot{r} + \frac{3}{2}\dot{r}^2 = \frac{1}{\rho}\left(P_g(r) + P_v - \frac{2\sigma}{r} - \frac{4\mu\dot{r}}{r} - P_\infty(t)\right). \tag{9}$$

In the Rayleigh–Plesset equation, r is the bubble radius, $P_g(r)$ is the gas pressure in the bubble, P_v is the saturated vapor pressure of the ink, and σ is the surface tension of the ink. The gas pressure is obtained from the polytropic relation assuming an isothermal bubble. The saturated vapor pressure is $P_v = 2400\ \text{N m}^{-2}$ and the surface tension is $\sigma = 0.028\ \text{N m}^{-1}$. Through the ambient pressure P_∞, which is obtained from the channel acoustics calculation, the channel acoustics are coupled to the bubble dynamics. A more extensive treatment of this model can be found in Ref. 14.

Electrically, the piezoactuator is a capacitor in parallel with a variable current source. The piezocurrent depends on the capacitance of the actuator C_a, the coupling coefficient α, and the pressure in the channel. The coupling coefficient relates the voltage over the piezo to the deformation of the channel and has also been used in the calculation of the channel acoustics. The time derivative of the charge expresses the relation between the actuator voltage and the piezocurrent in the time domain I_t,

$$I_t = \frac{dQ}{dt} = \left(\frac{\partial Q}{\partial U}\right)_P \frac{dU}{dt} + \left(\frac{\partial Q}{\partial P}\right)_U \frac{dP}{dt}. \tag{10}$$

Here Q is the total charge on the piezoactuator and U is the voltage over the piezoactuator. To calculate or interpret the piezocurrent, the isobaric capacitance and the relation between the channel pressure and current have to be determined. The isobaric capacitance is measured directly. The piezocurrent due to pressure fluctuations can be calculated from the thermodynamic fundamental equation of the actuator channel. The differential of the energy per unit length of channel is given by

$$de = PdA + Udq, \tag{11}$$

where q is the charge per unit length and e is the energy of the channel per unit length. Note that only the structure is considered in this section.

The analysis is simplified when the Legendre transform[27] with respect to pressure and actuator voltage is used, because the mechanical properties of the channel are

known in terms of the pressure and actuator voltage as independent parameters. The differential of the Legendre transform g (Gibbs the free energy per unit length) is

$$dg = de - d(AP) - d(Uq) = -AdP - qdU. \tag{12}$$

The isobaric capacitance is defined as the second derivative,

$$C_P \equiv \left(\frac{\partial Q}{\partial U}\right)_P = L_a\left(\frac{\partial q}{\partial U}\right)_P = -L_a\left(\frac{\partial^2 g}{\partial U^2}\right)_P, \tag{13}$$

where L_a is the actuator channel length and q is assumed to be constant. The coupling coefficient $(\partial Q/\partial P)_U$ is

$$\left(\frac{\partial Q}{\partial P}\right)_U = -L_a\left(\frac{\partial}{\partial P}\right)_U\left(\frac{\partial g}{\partial U}\right)_P = -L_a\left(\frac{\partial}{\partial U}\right)_P\left(\frac{\partial g}{\partial P}\right)_U = L_a\left(\frac{\partial A}{\partial U}\right)_P. \tag{14}$$

Combining Eqs. (1) and (14) yields the coupling coefficient

$$\alpha A L_a = \left(\frac{\partial Q}{\partial P}\right)_U. \tag{15}$$

Combining Eqs. (10), (13), and (15) yields the piezocurrent

$$I_t = C_p\frac{dU}{dt} + \alpha A L_a\frac{dP}{dt}. \tag{16}$$

In general, the pressure is a function of position. When the fluctuations are sufficiently slow for the system to come to rest locally, the piezocurrent can be obtained by integrating over the length of the actuator,

$$I_t = C_p\frac{dU}{dt} + \alpha A\int_0^{L_a}\frac{dP}{dt}dx. \tag{17}$$

This approximation is valid here, since the wavelength is much larger than the channel radius, ensuring that the system is in local equilibrium. The piezocurrent is now known in terms of the actuator voltage and the channel pressure.

From the Fourier transform of the pressure, the Fourier transform of the piezocurrent I_f can be calculated. Inserting the expression of the pressure into Eq. (17), applying the Fourier transform defined in Eq. (3), and dividing by $e^{i\omega t}$ yield an expression for the piezocurrent,

$$I_f = i\omega C_p U + \alpha A_j\frac{\omega}{k_j}(-P_{r,j}e^{-ikL_a} + P_{l,j}e^{ik_jL_a}) + \alpha_j A_j L_a i\omega P_s. \tag{18}$$

If the electric signal source were an ideal voltage source, the voltage over the actuator would now be prescribed and the electrical resistance would vanish. The piezocurrent would be determined and measured as an indication of the acoustics in the channel.

In reality, however, the signal generator is not an ideal voltage source but has an output impedance R_p. Therefore the voltage over the piezoactuator is not imposed but is obtained as a part of the solution. The symbols that refer to electric properties of the measurement system are clarified in Fig. 5. The piezovoltage is the sum of the actuation pulse U_a and the voltage over the output impedance of the signal generator and the connections,

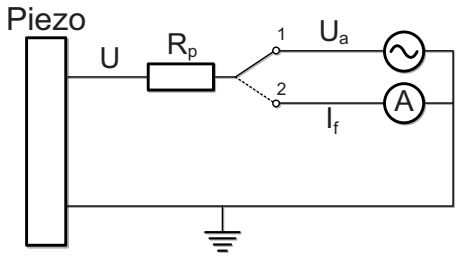


FIG. 5. The simplified measurement circuit: The switch changes between 1, actuation, and 2, when the piezo is used as hydrophone. The output impedance, R_p, is in reality distributed throughout the system. It consists of resistance at connections, in wires, in the ampere meter, and in the voltage source. The jet pulse U_a differs from the voltage U over the piezoelectrodes due to this resistance. In reality, the voltage source consists of a number of linked devices: an arbitrary waveform generator, a switchboard, and amplifiers.

$$U = U_a - I_f R_p. \tag{19}$$

When Eq. (19) is inserted into Eq. (18), an expression for the piezocurrent with a nonideal voltage source is obtained,

$$I_f = i\omega C_p(U_a - I_f R_p) + \alpha A_j\frac{\omega}{k_j}(-P_{r,j}e^{-ik_jL_a} + P_{l,j}e^{ik_jL_a}) + \alpha A_j L_a i\omega P_s. \tag{20}$$

Upon rearranging, the piezocurrent for a finite output resistance is obtained as

$$I_f = \frac{1}{1 + i\omega C_p R_p}\left(i\omega C_p U + \alpha_j A_j\frac{\omega}{k_j}(-P_{r,j}e^{-ik_jL_a} + P_{l,j}e^{ik_jL_a}) + \alpha_j A_j L_a i\omega P_s\right). \tag{21}$$

This expression shows that a finite output resistance acts as a low-pass filter with a cutoff frequency of $\omega_c = 1/C_p R_p$. Since the order of magnitude of the output impedance is typically $R_p = 100\ \Omega$ and the capacitance of the piezoactuator is about 1 nF, the cutoff frequency is typically $\omega_c = 10$ MHz. The order of magnitude of the resonance frequencies of the printhead is 100 kHz, which is much smaller. Therefore, the output impedance can be neglected.

The coupling coefficient α_j and the wall flexibility β_j can be determined by comparison of the measured and calculated piezocurrents. Modifying the coupling coefficient changes the magnitude of the measured signal, but not its shape. So when the correct value of α_j is used in the model, the amplitudes of the measured and calculated piezocurrents are equal. The wall flexibility changes the resonance frequencies of the channel. Thus, when the correct value of β_j is used, the frequencies that are present in the calculated piezocurrent match those in the measured signal. These conditions were used to determine both parameters. Now that these parameters have been determined, the current from a printhead with a bubble can be modeled and compared with the experiment (see Fig. 6).

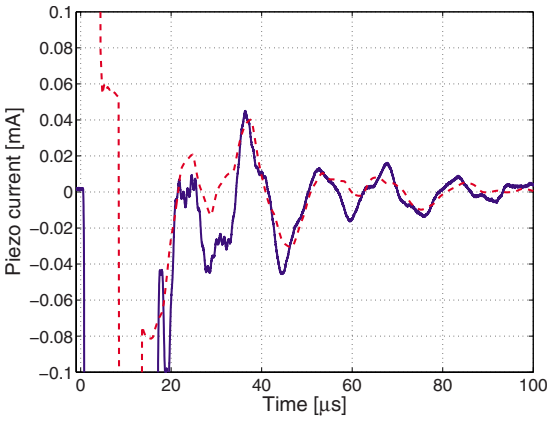


FIG. 6. (Color online) The measured (solid line) and calculated piezocurrents (dashed line). Both amplitude and frequency match which indicates that both the wall flexibility and the piezoelectric coupling coefficient are chosen correctly. The amplitude deviation in the beginning of the signal is probably caused by the dielectric relaxation of the piezo.

V. COMPARING THE MODEL WITH EXPERIMENTS

In order to compare the model with the experiment, it is convenient to single out the change in the piezocurrent due to the bubble. Therefore, the piezocurrent of the undisturbed nozzle I_0 is subtracted from the piezocurrent obtained when a bubble is entrapped $I(V_b)$. This gives the differential piezocurrent

$$\tilde{I}(V_b) = I(V_b) - I_0. \tag{22}$$

We will distinguish between the experimental differential current $\tilde{I}_e$, with a corresponding optical measured bubble volume V_e, and the differential current resulting from the model $\tilde{I}_m(V_m)$, where V_m is the volume of the bubble assumed in the calculation. The undisturbed piezocurrent is obtained experimentally by measuring the piezocurrent in the absence of an entrained bubble. With the model, the undisturbed current can be obtained by setting the bubble volume to zero. Figure 7 shows examples of experimentally obtained differential

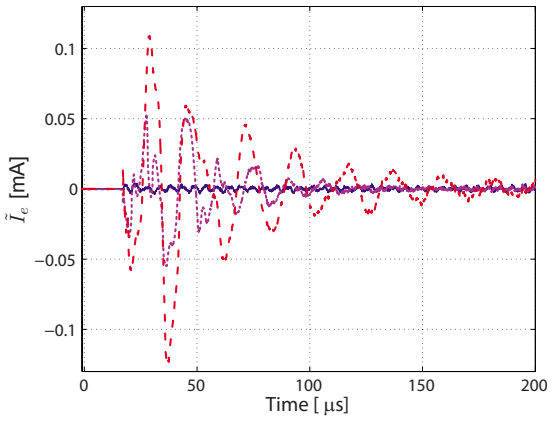


FIG. 7. (Color online) Experimentally obtained differential piezocurrents $\tilde{I}_e = I_e - I_0$. The solid line shows the signal of an undisturbed channel, the dotted line shows the signal when a bubble of 5 pl is entrapped, and the dashed line shows the signal when a bubble of 81 pl is entrapped. The signal from the undisturbed channel shows the magnitude of noise in the measurements. Obviously, in the absence of noise, the differential piezocurrent of the undisturbed channel would have vanished throughout.

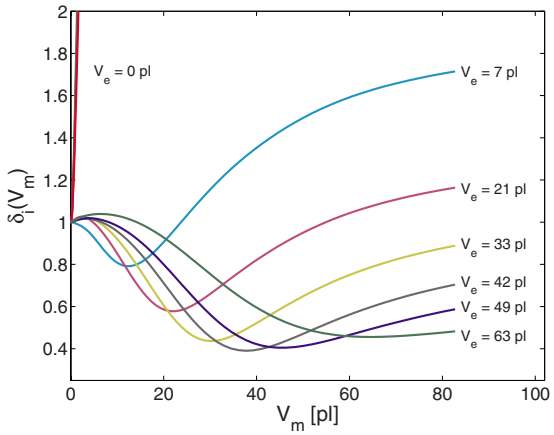


FIG. 8. (Color online) Norm of the difference between the measured and calculated disturbances. This function of the bubble volume has a distinct minimum where the agreement between the model and the experiment is the highest. The graphs where no bubble is present rise sharply from a value of $\delta_i(V_m)=1$ at zero bubble volume. The optically found bubble volumes are shown on the right of the curves; it agrees with the position of the minimum, revealing the success of the employed model.

currents $\tilde{I}_e$. This figure illustrates again the pronounced change in the piezocurrent when an air bubble is present, compared to the current of an undisturbed channel. Moreover, it shows that even for very small bubbles, the change in the piezocurrent is still significant.

The difference between the measured and calculated piezocurrents can be expressed as $\delta_i(V_m)$, the relative norm of the difference, defined as

$$\delta_i(V_m) = \frac{\|\tilde{I}_e - \tilde{I}_m(V_m)\|}{\|\tilde{I}_e\|}. \tag{23}$$

Here the L_2 norm is used, which is defined as

$$\|f(t)\|_2 \equiv \sqrt{\frac{1}{T}\int_0^T |f(t)|^2 dt}. \tag{24}$$

The norm of the difference is nondimensionalized using the norm of the measured differential current. In case of a bubble inside the channel, the value of $\delta_i(V_m)$ depends on the bubble volume V_m that is assumed in the calculation. The value of $\delta_i(V_m)$ is close to zero when the differential current of the model matches the differential current of the experiment. Note that $\delta_i(V_m)$ is a positive definite function of the bubble volume that is assumed in the calculation. Therefore, when $\delta_i(V_m)$ reaches a minimum, the match between model and experiment should be optimal. The value V_m for which this minimum is reached should then correspond to the measured bubble volume. In Fig. 8, the relative norm of the difference is shown for eight measured piezocurrents as a function of the assumed bubble volume. The functions are smooth and well behaved, which facilitates the search for their minimum. In the domain used in the calculation, only a single minimum is found for $\delta_i(V_m)$. To illustrate the agreement between the model and experiment at this minimum, Fig. 9 shows the differential piezocurrent of a measurement with its modeled counterpart. In this example, the optically obtained bubble volume was 81 pl. By inserting the corresponding piezocur-

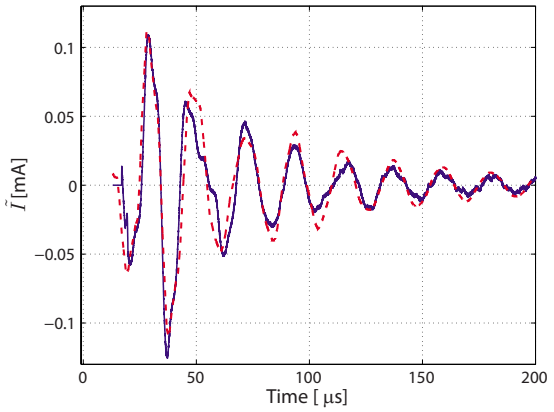


FIG. 9. (Color online) The calculated differential piezocurrent $\tilde{I}_m(V_m)$ (dashed) is compared with the experimentally obtained differential piezocurrent $\tilde{I}_e(V_e)$ (solid). For these currents, the model finds a bubble volume of 86 pl, which is close to the value of 81 pl that was measured optically.

rent into the model, the minimum in $\delta_i(V_m)$ was found for a bubble of 86 pl. As can be seen in Fig. 9, the calculated piezocurrent closely resembles the measured piezocurrent in both frequency and amplitude.

The quality of the model becomes even more convincing in Fig. 10, where V_e, gathered during the bubble dissolution process, is compared with V_m calculated by the model. For both methods, the absolute error is given by the colored area. The absolute error in the optically obtained bubble volume increases with the bubble volume. This originates from the measurement method, where the radius is extracted from the images with an accuracy of a few pixels. The absolute error is about 0.9 μm, independent of the bubble size itself. As the bubble volume is $V_e=\frac{4}{3}\pi r_e^3$, where r_e is the bubble radius, the relative error in the bubble volume is three times the relative error in the radius; $\Delta V_e/|V_e|=3\Delta r_e/|r_e|$. Correspondingly, the absolute error $\Delta V_e=(4\pi r_e^2)\Delta r_e$ is quadratic in the

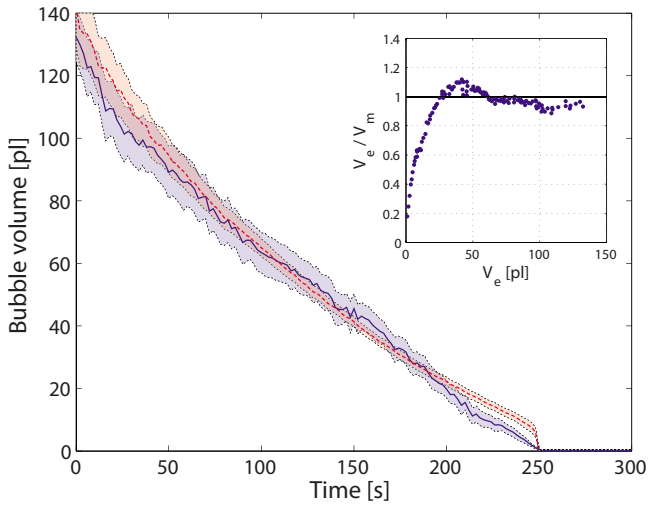


FIG. 10. (Color online) Acoustically measured bubble volume (V_m) is shown as a dotted line and the optically measured bubble volume (V_e) is shown as a solid line. The areas around the lines give the error margins in the results. In the inset, the ratio of the optically measured bubble volume over the acoustically measured bubble volume is shown. This illustrates that the relative error diverges for small bubbles. For larger bubbles, the relative error is less than 12%.

bubble radius. Note that the error in the optical bubble volume does not affect the error in the calculated result, as V_e is not a parameter of $\delta_i(V_m)$ but only the current I_e, which was measured simultaneously with V_e.

The error in the acoustic measurement ΔV_m is calculated from the minimum value in Fig. 8 by using

$$\Delta V_m = \frac{\|\tilde{I}_e - \tilde{I}_m(V_m)\|_2}{\frac{\partial}{\partial V_m}\|\tilde{I}_m(V_m)\|_1}. \tag{25}$$

Here the difference between the calculated piezocurrent and the measured piezocurrent is assumed to be Gaussian white noise. The derivative is evaluated by a finite difference approximation.

In the inset of Fig. 10, the ratio of the acoustically measured bubble volume over the optically measured bubble volume is shown. This illustrates that for bubbles above 20 pl, the relative error is less than 12%. For small bubble, the relative error diverges, and the acoustic measurement method becomes less accurate. This is attributed to nonlinear volume oscillations of the air bubble, which this linearized model cannot capture.

VI. SUMMARY AND OUTLOOK

A linear model is used to estimate the volume of a bubble in an inkjet channel. With this model, it is shown how a bubble influences the channel acoustics of an inkjet printhead. The linear approximation in this model is valid for bubbles that are larger than 20 pl. Small bubbles exhibit nonlinear behavior, which the model cannot capture. Therefore, the acoustic measurement method is less accurate in this regime. To overcome this problem, the method can be extended by solving the full nonlinear equations. The two-way coupling with the channel acoustics turns the Rayleigh–Plesset equation into a delay differential equation. This nonlinear equation can be solved numerically[28] at the cost of increased calculation time.

The model calculates the current through the actuator. By comparing the current with experimentally obtained currents, the model is able to accurately determine the bubble volume. In this way, an acoustic measurement method for the volume of entrapped air bubbles is obtained. This method was validated with optically measured bubble volumes. In addition, this shows that the linear regime of volume oscillations of an air bubble in an inkjet microchannel and the corresponding channel acoustics is well understood.

ACKNOWLEDGMENTS

This research could not be conducted without the help from Océ technologies B.V. and especially Jan Simons who designed the piezocurrent measurement system and Wilbert Classens and Ron Berkhout for their work on the glass channel section. This work is part of the research program of the Technology Foundation STW, which is financially supported by the "Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)" and by Océ Technologies B.V.

[1]T. G. Leighton, *The Acoustic Bubble* (Academic, London, 1994).
[2]C. E. Brennen, *Cavitation and Bubble Dynamics* (Oxford University Press, Oxford, 1995).
[3]M. S. Plesset and A. Prosperetti, "Bubble dynamics and cavitation," Annu. Rev. Fluid Mech. **9**, 145–185 (1977).
[4]M. P. Brenner, S. Hilgenfeldt, and D. Lohse, "Single-bubble sonoluminescence," Rev. Mod. Phys. **74**, 425–485 (2002).
[5]S. Qin and K. W. Ferrara, "Acoustic response of compliable microvessels containing ultrasound contrast agents," Phys. Med. Biol. **51**, 5065–5088 (2006).
[6]S. Qin and K. W. Ferrara, "The natural frequency of oscillation of ultrasound contrast agents in microvessels," Ultrasound Med. Biol. **33**, 1140–1148 (2007).
[7]C. F. Caskey, S. M. Stieger, S. Qin, P. A. Dayton, and K. W. Ferrara, "Direct observations of ultrasound microbubble contrast agent interaction with the microvessel wall," J. Acoust. Soc. Am. **122**, 1191–1200 (2007).
[8]V. S. Ajaev and G. M. Homsy, "Modeling shapes and dynamics of confined bubbles," Annu. Rev. Fluid Mech. **38**, 277–307 (2006).
[9]A. L. Klibanov, "Ultrasound contrast agents: Development of the field and current status," Top. Curr. Chem. **222**, 73–106 (2002).
[10]R. J. Dijkink, J. P. van der Dennen, C. D. Ohl, and A. Prosperetti, "The 'acoustic scallop': A bubble-powered actuator," J. Micromech. Microeng. **16**, 1653 (2006).
[11]K. S. F. Lew, E. Klaseboer, and B. C. Khoo, "A collapsing bubble-induced micropump: An experimental study," Sens. Actuators, A **133**, 161–172 (2007).
[12]J. de Jong, R. Jeurissen, H. Borel, M. van den Berg, M. Versluis, H. Wijshoff, A. Prosperetti, H. Reinten, and D. Lohse, "Entrapped air bubbles in piezo-driven inkjet printing: Their effect on droplet velocity," Phys. Fluids **18**, 121511 (2006).
[13]J. de Jong, G. de Bruin, H. Reinten, M. van den Berg, H. Wijshoff, M. Versluis, and D. Lohse, "Air entrapment in piezo-driven inkjet printheads," J. Acoust. Soc. Am. **120**, 1257–1265 (2006).
[14]R. Jeurissen, J. de Jong, H. Reinten, M. van den Berg, H. Wijshoff, M. Versluis, and D. Lohse, "Effect of an entrained air bubble on the acoustics of an ink channel," J. Acoust. Soc. Am. **123**, 2496–2505 (2008).
[15]B. Krasovitski and E. Kimmel, "Gas bubble pulsation in a semiconfined space subjected to ultrasound," J. Acoust. Soc. Am. **109**, 891–898 (2001).
[16]P. Zhong, Y. Zhou, and S. Zhu, "Dynamics of bubble oscillation in constrained media and mechanisms of vessel rupture in swl," Ultrasound Med. Biol. **27**, 119–134 (2001).
[17]J. Cui, M. Hamilton, P. Wilson, and E. Zabolotskaya, "Bubble pulsations between parallel plates," J. Acoust. Soc. Am. **119**, 2067–2072 (2006).
[18]H. Oguz and A. Prosperetti, "The natural frequency of oscillation of gas bubbles in tubes," J. Acoust. Soc. Am. **103**, 3301 (1998).
[19]E. Sassaroli and K. Hynynen, "Forced linear oscillations of microbubbles in blood capillaries," J. Acoust. Soc. Am. **115**, 3235–3243 (2004).
[20]E. Ory, H. Yuan, A. Prosperetti, S. Popinet, and S. Zaleski, "Growth and collapse of a vapor bubble in a narrow tube," Phys. Fluids **12**, 1268–1277 (2000).
[21]K. S. Kwon and W. Kim, "A waveform design method for high-speed inkjet printing based on self-sensing measurement," Sens. Actuators, A **140**, 75–83 (2007).
[22]J. F. Dijksman, "Hydrodynamics of small tubular pumps," J. Fluid Mech. **139**, 173–191 (1984).
[23]D. B. Bogy and F. E. Talke, "Experimental and theoretical study of wave propagation phenomena in drop-on-demand ink jet devices," IBM J. Res. Dev. **28**, 314–321 (1984).
[24]M. A. Groninger, P. G. M. Kruijt, H. Reinten, R. H. Schippers, and J. M. M. Simons, "A method of controlling an inkjet printhead, an inkjet printhead suitable for use of said method, and an inkjet printer comprising said printhead," European Patent No. EP 1 378 360 A1 (2003).
[25]J. R. Womersley, "Method for the calculation of velocity, rate of flow and viscous drag in arteries when the pressure gradient is known," J. Physiol. **127**, 553–563 (1955).
[26]T. Young, "Hydraulic investigations, subservient to an intended Croonian Lecture on the motion of the blood," Philos. Trans. R. Soc. London **98**, 164–186 (1808).
[27]H. Callen, *Thermodynamics and an Introduction to Thermostatistics* (Wiley, New York, 1985).
[28]L. F. Shampine and S. Thompson, "Solving DDEs in MATLAB," Appl. Numer. Math. **37**, 441–458 (2001).

Long range sound propagation over a sea surface

Karl Bolin[a] and Mathieu Boué
Marcus Wallenberg Laboratory, KTH, Stockholm 100 44, Sweden

Ilkka Karasalo
Marcus Wallenberg Laboratory, KTH, Stockholm 100 44, Sweden and Swedish Defense Research Agency, Stockholm 164 90, Sweden

(Received 26 February 2009; revised 21 August 2009; accepted 28 August 2009)

This paper describes methodology and results from a model-based analysis of data on sound transmission from controlled sound sources at sea to a 10-km distant shore. The data consist of registrations of sound transmission loss together with concurrently collected atmospheric data at the source and receiver locations. The purpose of the analysis is to assess the accuracy of methods for transmission loss prediction in which detailed data on the local geography and atmospheric conditions are used for computation of the sound field. The results indicate that such sound propagation predictions are accurate and reproduce observed variations in the sound level as function of time in a realistic way. The results further illustrate that the atmospheric model must include a description of turbulence effects to ensure predicted noise levels to remain realistically high during periods of sound shadow.
 [DOI: 10.1121/1.3238236]

PACS number(s): 43.28.Fp, 43.50.Vt [VEO] Pages: 2191–2197

I. INTRODUCTION

In the light of global warming, large-scale transition to renewable power sources is a worldwide challenge. One energy source that will play a significant role for this transition is wind turbine power. Until now most wind turbines are land based. However, large offshore farms are under construction or being planned all over the world, and the total worldwide capacity of this power source is projected to grow to approximately 50 GW by 2020.[1] Offshore wind turbines are often located in shallow waters near a coast and have therefore raised concerns for causing noise disturbances in adjacent coastal regions, often rural and recreational areas previously unaffected by community noise. Since atmospheric sound propagation is highly dependent upon the changing meteorological conditions noise levels may vary significantly with time. The need to optimize the power output from windfarms under strict constraints on noise pollution motivates an interest in techniques for accurate prediction of sound propagation from offshore wind turbines.

Measurement of long distance sound propagation over sea surfaces together with concurrent registration of meteorological data have been performed by Konishi and Tanioku[2] and Konishi.[3] However, the meteorological data were collected up to a few hundred meters height only, while knowledge of the atmospheric conditions (wind velocity, humidity, and temperature) further up in the atmosphere could severely influence the sound field.

This paper presents measurements of the transmission loss (TL) of sound propagated over sea to a 10-km distant receiver on land[4] and compares the experimental TL data with numerical predictions. The predictions are computed using the Green's function parabolic equation (GFPE) method[5,6] using atmospheric data from concurrent meteorological measurements at the source and the receiver sites. The purpose of this work is to assess the reliability of predictions of sound transmission with numerical models which use detailed knowledge of the meteorological and geographical conditions.

II. MEASUREMENTS

The measurements were conducted from the 15th to the 21st of June 2005 in the Kalmar strait and the island Öland in the Baltic Sea; see Fig. 1. This period was chosen because most annoyance from wind turbine noise is expected in the summer. Also, low level jets are a frequently occurring meteorological condition in the early summer season.[7] These conditions are characterized by local wind speed maxima at heights below approximately 1 km. Sound propagating in the downwind direction will then be trapped in a sound channel bounded upward by this maximum, leading to cylindrical spreading of the sound energy in the far field and thus increased sound levels compared to spherical spreading.

A. Acoustical measurements

1. Sources

Two sound sources were placed on the Utgrunden lighthouse located 9 km from shore (WGS84 coordinates N 56° 22.40′, E 16° 15.50′). These sources were mounted on the lighthouse roof at a height of 30 m with reference microphones located 1 m in front of respective source for recording the emitted signals; see Fig. 2.

The first source was a compressed-air-driven sound source (Kockums Sonics Supertyfon AT150/200 with Valve Unit TV 784). It produced a 10-s signal with average source level of 130 dB at 200 Hz. Both the 200-Hz fundamental

[a]Author to whom correspondence should be addressed. Electronic mail: kbolin@kth.se

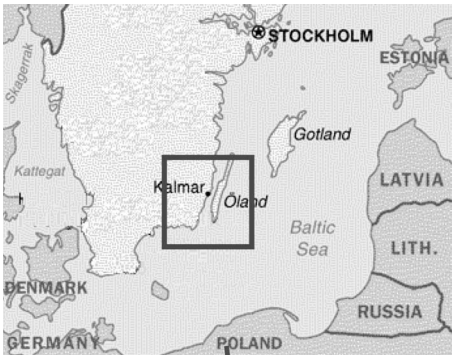


FIG. 1. Map of the Southern Baltic Sea with the Kalmar straight shown by the square.

tone and the first harmonic of this signal were used in the analysis. The second source consisted of a sound generator coupled to a loudspeaker and a 1.2-m-long resonator tube. It emitted a 1-min-long 80-Hz tone with a constant sound pressure level of 113 dB. Both sound sources were employed simultaneously.

The choice of source frequencies was motivated by the spectral characteristics of wind turbine noise, as illustrated in Fig. 3, showing third-octave averaged noise as function of frequency at 500-m distance from a modern 2-MW offshore wind turbine. The noise level is seen to decrease with frequency, in particular, for frequencies above approximately 500 Hz, a characteristic that would be further enhanced by the increasing sound attenuation at propagation to the longer distances considered here.

2. Receiver

The receiver site was on the island Öland, 750 m from shore in a very quiet residential area with ground altitude of 7 m above the sea (WGS84 coordinates N 56° 23.35′, E 16° 24.67′).

The receiver was a linear array of eight 0.5-in. microphones oriented parallel to the direction toward the source. The microphones were placed at 1.7-m height according to ISO 1996.[8] The distance between the microphones was set to $d=40$ cm, equal to half the wavelength at 400 Hz, to ensure directivity patterns free from grating lobes at all three frequencies. The signals were transmitted through a preamplifier to a UA100 analyzer and then processed in MATLAB as explained below. The signals $x(t)$ of the N microphones were added with their respective time delays τ, as shown by Eq. (1),

$$s(t) = \frac{1}{N}\sum_{n=1}^{N} w(n)x_n(t-\tau_n), \tag{1}$$

where $w(n)$ are the binomial coefficients $N!/n!(N-n)!$, $\tau_n = (n-1)c^{-1}d\cos\phi$ is the time delay of the nth microphone, x_n is the signal recorded by the nth microphone, $\phi = 0$ is the angle between the direction of propagation and the direction of the array, and c is the sound speed.

When atmospheric conditions were unfavorable the delay-and-sum beamforming in Eq. (1) had to be combined with a frequency tracking algorithm, as described in Ref. 4 (Chap. 4), to ensure detection of the signal against the background noise.

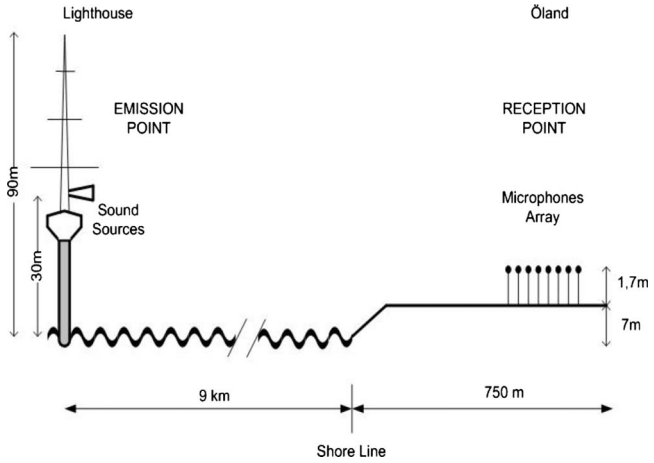


FIG. 2. In situ setup.

B. Meteorological measurements

1. Source site

At the source site the horizontal wind speed and the temperature were registered at heights of 38, 50, 65, 80, and 90 m by anemomenters and thermometers mounted on a meteorological mast attached to the lighthouse. The instruments recorded averages and, for the wind speed, standard deviations over 10-min time intervals.

2. Receiver site

Meteorological profiles at the receiver site were measured several times daily during day-time using radio probes and theodolite tracking of free flying balloons.[9] These measurements were performed by staff from the Department of Earth Sciences, Uppsala University. Wind velocity (horizontal components), humidity, and temperature were measured up to 3500-m height.

III. SOUND PROPAGATION MODEL

A sound propagation model suitable for the current experimental scenario must fulfill two principal requirements. First, the model must allow the atmospheric parameters to vary both in vertical and horizontal directions. Second, it must be able to handle propagation of sound into shadow

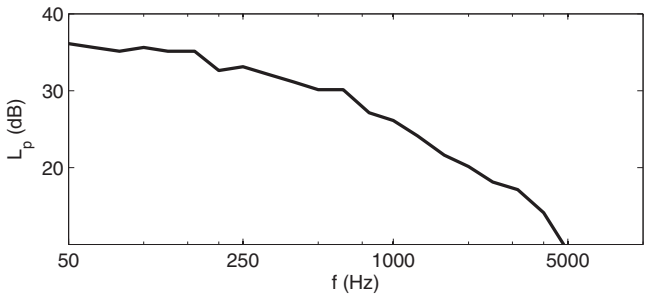


FIG. 3. Third octave band sound levels from a 500-m distant modern (2-MW) offshore wind turbine from Ref. 21.

zones (diffraction and scattering) since the atmospheric conditions were such that the receiver was in sound shadow during more than 60% of the trial week. Therefore, a parabolic equation (PE)-method was considered appropriate for the current application.

A. The GFPE method

The GFPE method was developed by Gilbert an Di[5,10] and later improved by Salomons.[6,11] The method is particularly designed for atmospheric sound propagation and can use considerably longer range-steps than conventional PE-methods. Because of its computational efficiency, the GFPE model was used in this study.

The method computes a two-dimensional field in the rz-plane where r is the radial distance from the source and z is the height. From the three-dimensional Helmholtz equation for the sound pressure p in cylindrical coordinates combined with a variable substitution $\phi = \exp(-ik_0 r) p r^{1/2}$ expressions (2) and (3) can be derived,[5,6]

$$\phi(r+\Delta r,z) = \exp\left(i\frac{\Delta r \delta k 2(z)}{2k_r}\right)\left[\frac{1}{2\pi}\int_{-\infty}^{\infty}(\Phi(r,k') + R(k')\Phi(r,-k'))\exp(i\Delta r(\sqrt{k_r 2 - k'^2} - k_r)) \times e^{ik'z}dk' + 2i\beta\Phi(r,\beta)\exp(i\Delta r(\sqrt{k_r 2 - \beta 2} - k_r))e^{-i\beta_z}\right], \quad (2)$$

where Δr is the horizontal step size, $k(z) = \omega / c(z)$ is the wave number, k_r is a reference wave number [$k_r = k_0 = k(0)$ in this paper[6]], $R(k') = (k' Z_g - k_r)/(k' Z_g + k_r))$ is the plane-wave reflection coefficient, Z_g is the normalized ground impedance, $\beta = k_r / Z_g$ is the surface-wave pole in the reflection coefficient, and $\Phi(r,k)$ is given by

$$\Phi(r,k) = \int_0^{\infty} \exp(-ikz')\phi(r,z')dz'. \quad (3)$$

Equations (2) and (3) combined constitute the fundamental step in the GFPE-algorithm. In our implementation the integrals are computed by the midpoint rule, and the propagation factor, first term on the right-hand side of Eq. (2), is substituted by

$$\exp(i\Delta r[k(z) - k_r]). \quad (4)$$

The starting sound pressure profile is a Gaussian function of height z at range $r=0$,

$$\phi(0,z) = \left[e^{-k_0 2(z-z_s)2} + \frac{Z_g - 1}{Z_g + 1} e^{-k_0 2(z+z_s)2}\right], \quad (5)$$

where z_s is the source height.

B. GFPE method parameters

The parameters of the GFPE method were guided by suggestions in Refs. 5, 6, and 11. Thus, the horizontal and vertical step sizes Δr and Δz depended on the wavelength at the ground λ. These were set to $\Delta r = 10\lambda$ and $\Delta z = 0.1\lambda$ in accordance with recommendations from Ref. 6. To suppress spurious reflections from the upper boundary of the computational domain, an artificial absorption layer with thickness of 75λ was imposed with an absorption parameter A calculated according to Ref. 6. Sound attenuation coefficients were calculated in accordance to ISO 9613-1.[12]

Our implementation of the GFPE model was validated by comparisons with the reference cases considered in Ref. 5. The predicted TL showed perfect agreement with the results in Ref. 5.

C. Surface conditions

At calm seas the acoustic impedance at the water surface is very high. Consequently, the surface would be almost totally reflecting. However, under normal conditions the rough sea surface caused by wind-driven waves induces random scattering of the reflected sound. This scattering could be represented approximatively by a modification of the surface impedance. In the model predictions shown below, the Delany and Bazley[13] surface impedance model was used for both the sea and the ground surfaces. The impedance is then determined by the sound frequency and a flow resistivity parameter, which for the sea surface chosen to 3×10^5 cgs. The alternative to use boss theory[14] for modeling an equivalent impedance of a surface with periodically occurring obstacles was considered not to be applicable since its validity, for prevalent wavelengths of the sea waves at the site, is limited to frequencies below 40 Hz, as shown in Ref. 15.

For the ground surface the flow resistivity parameter was allowed to be frequency-dependent, with values chosen to provide a fit to the low TL values in the experimental data. The flow resistivity values obtained in this way were 200 cgs rayls for 200 and 400 Hz and 4×10^5 cgs rayls for 80 Hz. The value of 200 cgs rayls corresponds to grass in a rough pasture,[16] which is fairly consistent with the ground materials at the beach composed of pebbles and gravel at the shoreline followed by pasture land further up. The 80-Hz value is unrealistic as it corresponds to a nearly perfectly reflecting ground surface. A detailed analysis of the causes of this discrepancy was not considered meaningful, however, since modeling of the sea-ground surface as a locally reacting smooth boundary is of course only a crude approximation of the actual surface interaction.

D. Meteorological parameters

Meteorological input to the GFPE model was both the wind balloon data (horizontal wind velocity), radio balloon [relative humidity (rh), atmospheric pressure p, and temperature T], and the anemometers on the mast (standard deviation of wind speed). The wind balloon measurements were used as meteorological parameters $(\mathbf{U}(z), \mathrm{rh}, T, p)$ for the laminary atmosphere while the data from the mast were used to estimate the turbulence intensity (standard deviation of wind speed and temperature). Linear interpolation was used between measurement points in the vertical direction as well as in time. The range-independent laminar effective sound velocity fields were calculated at each hour as the sum of the horizontal laminar wind velocity component in the direction toward the receiver and the sound speed in a non-moving

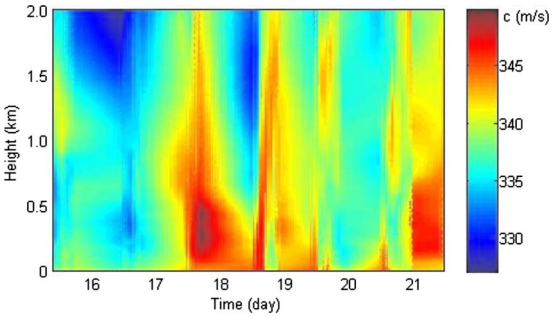


FIG. 4. (Color online) Sound speed profiles (m/s) during the measurement period.

atmosphere as function of rh, T, and p according to Ref. 17. Figure 4 shows the laminar range-independent effective sound speed as function of height and time throughout the experimental period. Local wind maxima, i.e., low level jets, occur in the afternoons of June 17th and 21st. A sudden shift in the profile can be seen at the beginning of June 21st.

E. Turbulence

Effects of turbulent wind and temperature fields were included in the GFPE model following the approach outlined in Ref. 11. Thus, the turbulent components of these fields are modeled as homogeneous random fields with von Karman spectra. The effect of such turbulence on the GFPE solution is represented by including a random z-dependent phase factor in the GFPE propagator, without requiring explicit computation of realizations of the fields (Appendixes I and J in Ref. 11). The two-dimensional von Karman spectral density $F(k_x,k_z)$ [Eq. (1.53) in Ref. 11] is

$$F(k_x,k_z) = \frac{A}{(k^2+K_0^2)^{8/6}}\left(\frac{\Gamma(1/2)\Gamma(8/6)}{\Gamma(11/6)}\frac{C_T^2}{4T_0^2} + \left[\frac{\Gamma(3/2)\Gamma(8/6)}{\Gamma(17/6)} + \frac{k_z^2}{k_z^2+K_0^2}\frac{\Gamma(1/2)\Gamma(14/6)}{\Gamma(17/6)}\right]\frac{11C_v^2}{c_0^2}\right), \quad (6)$$

where $A \approx 0.0330$, Γ is the gamma function, $k^2 = k_x^2 + k_z^2$, C_T^2 and C_v^2 are structure parameters depending on temperature and velocity, T_0 is the average temperature from the mast measurements, and c_0 is the sound speed at ground level at the receiver site. The lower boundary of the turbulent wave number K_0 is set to 10 m^{-1}, and 100 uniformly distributed wave numbers were used to discretize the spectrum. Temperature fluctuations were unfortunately not measured and their value was modeled according to Ref. 11 with the structure parameter, C_T^2, set to $10^{-6}T_0^2$. The standard deviation σ_u of the wind speed and T_0 as function of time during the experimental week are shown in Fig. 5. The magnitude of the von Karman spectrum is governed by C_T^2 and C_v^2 which is coupled to σ_u, as shown in Ref. 18 [Eq. (7.111)] for the wind speed component

$$\sigma_u^2 = \frac{\Gamma^2(1/3)}{\pi 2^{4/3}\sqrt{3}K_0^{2/3}}C_v^2. \quad (7)$$

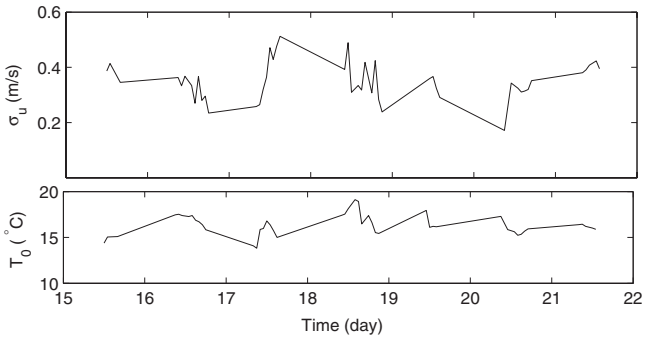
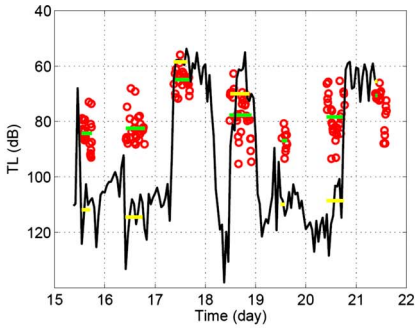
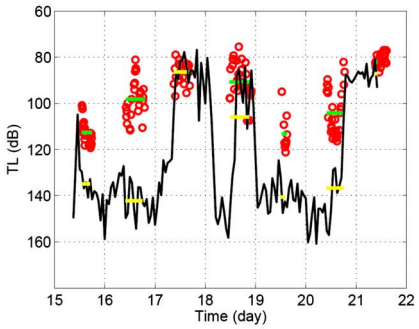


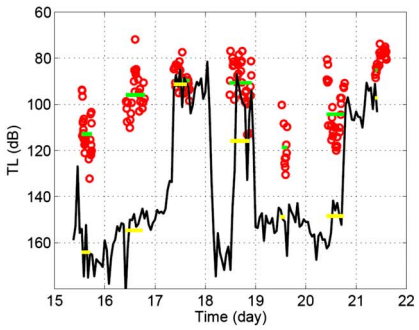
FIG. 5. Standard deviation of wind speed (upper graph) and temperature (lower graph) at the source point during the measurement period.



(a) Transmission loss at 80 Hz is shown as a function of time.



(b) Transmission loss at 200 Hz is shown as a function of time.



(c) Transmission loss at 400 Hz is shown as a function of time.

FIG. 6. (Color online) Measured (○) and predicted (—) TLs for the laminar calculation. Daily averages of measured and predicted TLs are shown as horizontal lines.

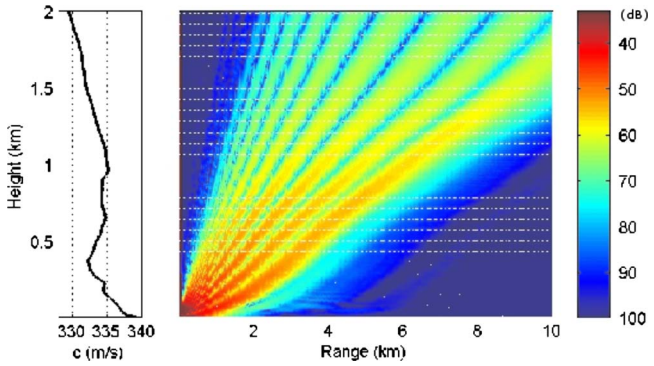


FIG. 7. (Color online) Effective sound speed (left) and model-predicted transmission loss (right) are shown for the 80-Hz case at 12 am the 16th June with laminar wind field.

According to this turbulence model the transmission loss to the receiver is a stochastic variable. The statistics of the transmission loss were determined by carrying out 50 Monte Carlo runs for each frequency at every hour during the measurement period.

IV. RESULTS

In this section the numerical predictions of the transmission loss are presented and compared to the experimental data. The results are shown in Figs. 6 and 10 using an atmospheric model without and with turbulence, respectively.

A. Laminar atmospheric model

The black curves in Fig. 6 show the calculated TL as function of time during the week. The daily average values during measurement periods are shown as horizontal lines. Measured TL values are shown as dots, and daily average values are shown as horizontal lines. It can be clearly seen that low TLs show good agreement with the measured TLs. Whereas, high TL values are severely overestimated by the predictions. The high TL values occur when the sound speed monotonically or nearly monotonically decreases with height. The emitted sound is then refracted upward and shadow zone occurs at the receiver location. A typical occasion showing this condition is shown in Fig. 7. The low TL values occur when the sound speed has a local maximum at relatively low height, causing the sound to be trapped within a channel below the local wind maximum. Such meteorological conditions occurred, for instance, in the afternoon of June 17, as can be seen in Fig. 8.

TABLE I. Average over all days of the daily normalized differences between predicted and measured TLs. The daily normalized difference is defined as the average over 1 day of the difference between measured and predicted TLs divided by the standard deviation that day of the measured TL.

Frequency (Hz)	Laminar atmosphere	Turbulent atmosphere
80	4.28	1.28
200	3.23	1.44
400	4.29	1.26

The *Dn* values shown in Table I are normalized differences between measured and predicted TLs. These are defined as the average of the differences between measured and predicted TLs divided by the standard deviation of the TL measurements at each day.

B. Turbulent atmospheric model

Turbulence in the atmospheric wind and temperature fields introduce random inhomogeneities in the sound speed. The inhomogeneities induce some random scattering of the sound, leading to increased leakage of sound into shadow zones, as noted by McBride *et al.*[19] and L'Esprance and Daigle[20] for shorter propagation distances. This effect is illustrated in Fig. 9, showing the sound field obtained with the same sound speed profile as in Fig. 7 but with effects of turbulence included. The difference between the sound fields is mainly that the shadow zone is less pronounced in a turbulent atmosphere.

In Fig. 10 the predicted TL as function of time including effects of turbulence is shown. The thick black curves show the average value of the TL from the Monte Carlo process, and the thinner black curves surrounding these show the interval of the standard deviations. Other symbols in the figures are defined as in Fig. 6. By comparing Figs. 6 and 10 it can be seen that the most prominent effect of turbulence on the predictions is a significant decrease in the TL during periods of sound shadow at the receiver. This could be explained by the random scattering of sound caused by the turbulence leading to increased sound levels in the shadow

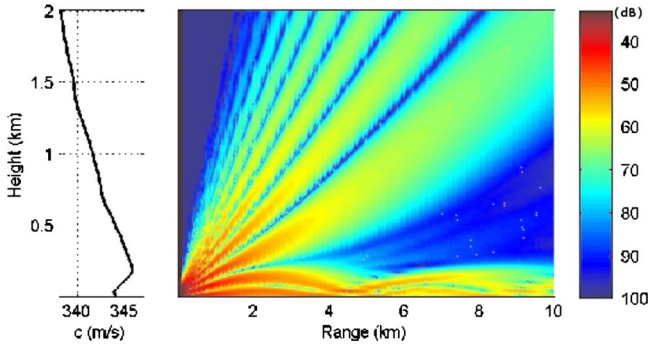


FIG. 8. (Color online) Effective sound speed (left) and model-predicted transmission loss (right) are shown for the 80-Hz case at 8 pm the 17th June with laminar wind field.

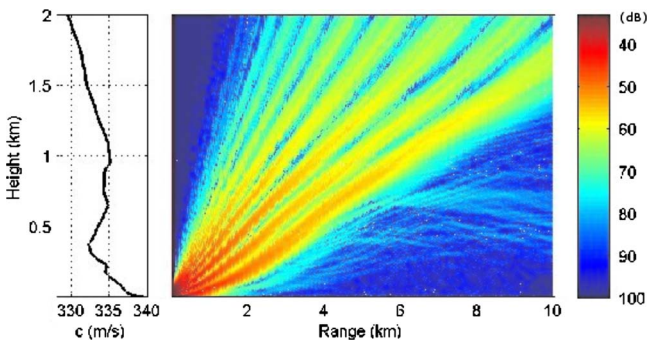
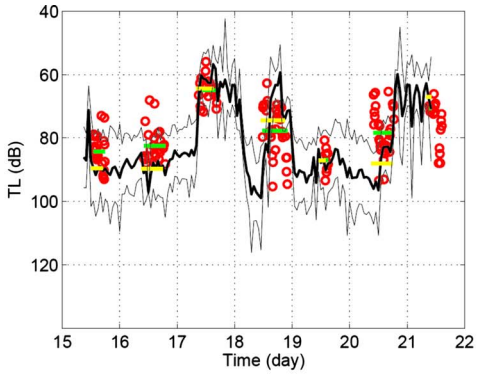
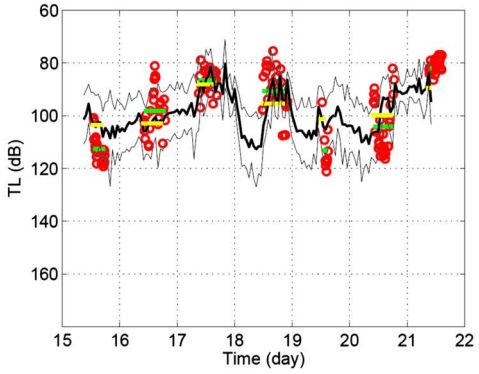


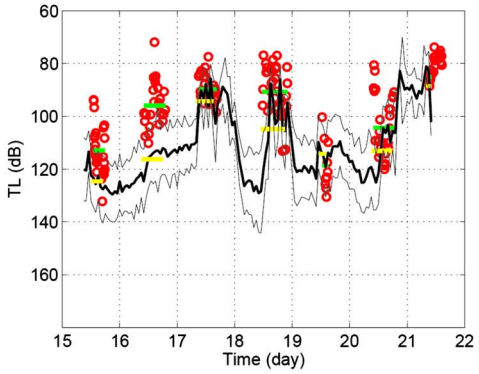
FIG. 9. (Color online) Effective sound speed (left) and model-predicted transmission loss (right) are shown for the 80-Hz case at 12 am the 16th June with turbulent wind field.



(a) Transmission loss at 80 Hz is shown as a function of time.



(b) Transmission loss at 200 Hz is shown as a function of time.



(c) Transmission loss at 400 Hz is shown as a function of time.

FIG. 10. (Color online) Measured (○) and predicted (—) TL for the turbulent calculation. Daily averages of measured and predicted TL are shown as horizontal lines.

zones. As seen in Table I the agreement between the predicted and the experimentally observed TL has thereby improved significantly.

V. CONCLUSIONS

The results support that sound propagation modeling including effects of detailed meteorological data can be used for reliable predictions of TLs. In particular, the predicted TL remains reasonably accurate under varying meteorological conditions and follows the variations observed in the TL measurements in a realistic way. The results further indicate that the sound propagation model must include effects of turbulence in the atmosphere for accurate predictions of the TL into shadow zones.

ACKNOWLEDGMENTS

The authors wish to thank Associate Professor Hans Berggren and M. Sc. Karin Törnblom, Department of Earth Sciences, Uppsala University, for measuring the atmospheric conditions. The authors wish to thank Professor Sten Ljunggren, Associate Professor Hans Bersgström, and Civ. Ing Hans Ohlsson for helpful comments and discussions. The Swedish Energy Authority is acknowledged for its financial support through the VINDFORSK II program.

[1]European Wind Energy Association, "EWEA's response to the European Commission's green paper" (COM (2006) 275 Final), 2007.
[2]K. Konishi, Y. Tanioku, and Z. Maekawa, "Long time measurement of long range sound propagation over an ocean surface," Appl. Acoust. **61**, 149–172 (2000).
[3]K. Konishi and Z. Maekawa, "Interpretation of long term data measured continuously on long range propagation over sea surfaces," Appl. Acoust. **62**, 1183–1210 (2001).
[4]M. Boué, "Long-range outdoor sound propagation over sea, applications to wind turbine noise," Technical Report No. TRITA-AVE 2007:22, KTH, Stockholm, Sweden, 2007.
[5]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).
[6]E. M. Salomons, "Improved Greens function parabolic equation method for atmospheric sound propagation," J. Acoust. Soc. Am. **104**, 100–111 (1998).
[7]A. K. Blackadar, "Boundary layer wind maxima and their significance for the growth of nocturnal inversions," Bull. Am. Meteorol. Soc. **38**, 282–290 (1957).
[8]International Organization for Standardization, Geneva, Switzerland, ISO1996: Acoustics—Description and measurement of environmental noise (1986).
[9]K. Törnblom, "Thermally driven wind modification in coastal areas and its influence on sound propagation with application to wind power," Technical Report No. DIVA-86480, Department of Earth Sciences, Uppsala University, Uppsala, Sweden, 2006.
[10]X. Di and K. E. Gilbert, in Proceedings of the 5th International Symposium on Long Range Sound Propagation, Milton Keynes, London 24–26 May 1992, pp. 128–146.
[11]E. M. Salomons, *Computational Atmospheric Acoustics* (Kluwer Academic, Dordrecht, 2002).
[12]International Organization for Standardization, Geneva, Switzerland, ISO 9613-1: Attenuation of sound during propagation outdoors—Part 1: Atmospheric absorption (1995).
[13]M. E. Delany and E. N. Bazley, "Acoustical properties of fibrous absorbant materials," Appl. Acoust. **3**, 105–116 (1970).
[14]P. Boulanger, K. Attenborough, S. Taherzadeh, T. Waters-Fuller, and K. M. Li, "Ground effect over hard rough surfaces," J. Acoust. Soc. Am. **104**, 1474–1482 (1998).
[15]L. Johansson, "Sound propagation around off-shore wind turbines—Long-range parabolic equation calculations for Baltic sea conditions," Technical Report KTH-BYT 192, KTH/Building Sciences, Stockholm, Sweden, 2003.
[16]T. F. W. Embleton, J. E. Piercy, and G. A. Daigle, "Effective flow resistivity of ground surfaces determined by acoustical measurements," J. Acoust. Soc. Am. **74**, 1239–1244 (1983).
[17]International Organization for Standardization, Geneva, Switzerland, ISO 9613-2: Attenuation of sound during propagation outdoors—Part 2: General method of calculation (1996).
[18]V. E. Ostashev, *Acoustics in Moving Inhomogeneous Media* (E. and F. N.

Spon, New York, NY, 1997).
[19]W. E. McBride, H. E. Bass, R. Raspet, and K. E. Gilbert, "Scattering of sound by atmospheric turbulence: Predictions in a refractive shadow zone," J. Acoust. Soc. Am. **91**, 1336–1340 (1992).
[20]A. L'Esprance, Y. Gabillet, and G. A. Daigle, "Outdoor sound propagation in the presence of atmospheric turbulence: Experiments and theoretical analysis with the fast field program algorithm," J. Acoust. Soc. Am. **98**, 570–579 (1995).
[21]Danish Environmental Protection Agency, Noise from offshore wind turbines (2005), Environmental Project No. 1016.

Acoustic intensity-based method for sound radiations in a uniform flow

Chao Yu[a)]
Department of Mechanical Engineering, Michigan State University, East Lansing, Michigan 48824

Zhengfang Zhou
Department of Mathematics, Michigan State University, East Lansing, Michigan 48824

Mei Zhuang[b)]
Department of Mechanical Engineering, Michigan State University, East Lansing, Michigan 48824

(Received 22 April 2009; revised 5 August 2009; accepted 18 August 2009)

An acoustic intensity-based method (AIBM) is extended and verified for predicting sound radiation in a subsonic uniform flow. The method assumes that the acoustic propagation is governed by the modified Helmholtz equation on and outside of a control surface, which encloses all the noise sources and nonlinear effects. With acoustic pressure derivative and its co-located acoustic pressure as input from an open control surface, the unique solution of the modified Helmholtz equation is obtained by solving the least squares problem. The AIBM is coupled with near-field Computational Fluid Dynamics (CFD)/Computational Aeroacoustics (CAA) methods to predict sound radiation of model aeroacoustic problems. The effectiveness of this hybrid approach has been demonstrated by examples of both tonal and broadband noise. Since the AIBM method is stable and accurate based on the input acoustic data from an open surface in a radiated field, it is therefore advantageous for the far-field prediction of aerodynamics noise propagation when an acoustic input from a closed control surface, like the Ffowcs Williams–Hawkings surface, is not available [Philos. Trans. R. Soc. London, Ser. A **264**, 321–342 (1969)].
 [DOI: 10.1121/1.3224764]

PACS number(s): 43.28.Ra, 43.20.Bi, 43.20.Fn, 43.20.El [SFW] Pages: 2198–2205

NOMENCLATURE

A_0 = Acoustic source strength
c = Speed of sound
H_n = nth order Hankel function of the second kind
i = $\sqrt{-1}$
k = Wave number, ω/c
M = Number of measurements (inputs)
M_a = Mach number
$\mathbf{n}$ = Unit normal vector, (n_x, n_y)
N = Terms of approximation in the asymptotic formulations
p = Acoustic pressure in the time domain
P = Acoustic pressure in the frequency domain
r, θ = Polar coordinates
$\hat{r}, \hat{\theta}$ = Modified polar coordinates
$\mathbf{U}$ = Free stream velocity vector, (U, V)
x, y = Cartesian coordinates
$\hat{x}, \hat{y}$ = Modified Cartesian coordinates
t = Time

Greek

α = Angle of attack
β = $\sqrt{1-M_a^2}$
Φ = Velocity potential function
γ = Angle between the two measurement segments
Γ = Spherical boundary containing all the acoustic sources
Γ_1 = Partial boundary of Γ
ρ_0 = Free stream density
ω = Angular frequency

I. INTRODUCTION

To effectively reduce aircraft noise, it is crucial to understand characteristics of aerodynamic noise sources. These characteristics, however, are not known analytically for aeroacoustic problems of practical significance. Various methods have been developed for characterizing sound sources and predicting their radiated acoustic field. For the far-field acoustic prediction, several methods (e.g., Refs. 1–4) have been applied in the past, with the Ffowcs Williams–Hawkings (FW-H) integral method[1] as the most common one. For the integral methods, the calculation of the acoustic pressure at each far-field location requires surface integrations around a closed control surface. This is computation intensive, especially for a practical three-dimensional problem. The acoustic intensity-based method (AIBM),[4] like the Helmholtz Equation Least-Squares (HELS) method in

[a)]Present address: Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, Florida 32611. Author to whom correspondence should be addressed. Electronic mail: yuchao@ufl.edu
[b)]Present address: Department of Aerospace Engineering, The Ohio State University, Columbus, Ohio 43210.

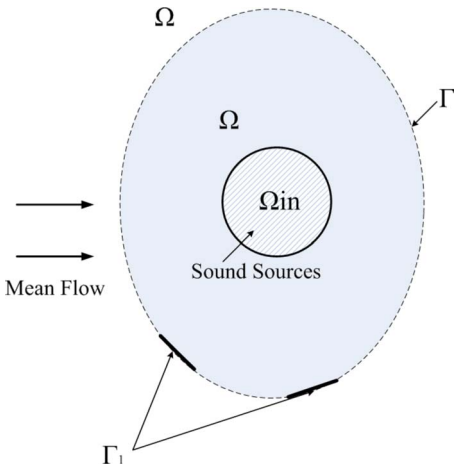


FIG. 1. (Color online) Schematic diagram of sound propagation field and locations of acoustic measurements.

inverse acoustics,[5] considers that the general solution for the Helmholtz equation is approximated by a linear combination of basis functions. After determination of the coefficients by matching the assumed form of the solution to the input acoustic data (measured or calculated), the far-field acoustic pressure can be calculated directly from the analytical expression. The method is, therefore, very efficient. In addition, because of the inclusion of pressure derivative in the input data, the method has less dependency on the completeness of the acoustic input data from a control surface.[6–8] The method has been verified in a stationary flow.[4] Since most aeroacoustic problems involve sound propagation in a flow field, in this paper, the method is extended to include sound radiation in a uniform subsonic flow. Furthermore the effectiveness of AIBM when coupled with near-field CFD/CAA methods is demonstrated by test examples.

This paper is organized as follows. Section II describes the mathematical formulation and numerical implementation of the AIBM in a uniform flow. In Sec. III, an example of acoustic radiation from a monopole source in a uniform flow is solved for the verification of the formulation. Two additional cases, one involving the aerodynamic noise radiation of flow around a NACA airfoil and the other considering a broadband sound scattering problem, are also studied in that section. The results demonstrate the effectiveness of the AIBM as a far-field prediction method by the coupling of the method with near-field CFD/CAA methods. Furthermore, the solutions from the AIBM are also compared with that from the FW-H method. The conclusions are drawn in Sec. IV.

II. MATHEMATICAL FORMULATIONS

In this section, the formulations of the AIBM (Ref. 4) are extended to include sound propagations in a uniform flow. Without loss of generality, we assume the uniform mean velocity is in the x-direction, $\mathbf{U}=U\hat{i}$. Let Ω_{in} be a bounded domain in R^2 containing all acoustic sources (see Fig. 1), and c be the speed of sound, it is well-known that the acoustic pressure $p(x,y,t)$ is governed by the following homogeneous wave equation:

$$\nabla^2 p - \frac{1}{c^2}(\partial_t + \mathbf{U}\cdot\nabla)^2 p = 0. \tag{1}$$

With Fourier transformation, the wave equation can be transformed to the modified Helmholtz equation. Assume that $p = e^{i\omega t}P(x,y)$ with angular frequency ω, $P(x,y)$ satisfies

$$\nabla^2 P - M_a{}^2 P_{xx} - 2ikM_a P_x + k^2 P = 0, \tag{2}$$

where $k=\omega/c$ is the wave number. $M_a=U/c$ represents the Mach number, which is assumed to be less than 1 in the current study. In order to get the solution of Eq. (2), the equation is converted to the standard Helmholtz equation. By employing the Prandtl–Glauert transformation, and setting $W(\hat{x},\hat{y})=P(x,y)$ with $(\hat{x},\hat{y})=(x/\beta,y)$ and $\beta=\sqrt{1-M_a{}^2}$, Eq. (2) can be rewritten as

$$\nabla^2 W - \frac{2ikM_a}{\beta}W_{\hat{x}} + k^2 W = 0. \tag{3}$$

To eliminate the first order term $W_{\hat{x}}$, another function $S(\hat{x},\hat{y})$ is introduced and defined as

$$S(\hat{x},\hat{y}) = \exp(-ikM\hat{x}\beta^{-1})W(\hat{x},\hat{y}). \tag{4}$$

The equation for S is therefore expressed as

$$\nabla^2 S + \frac{k^2}{\beta^2}S = 0. \tag{5}$$

In terms of the polar coordinates for $\hat{x}\hat{y}$-plane, $\hat{x}=\hat{r}\cos\hat{\theta}$, $\hat{y}=\hat{r}\sin\hat{\theta}$ and $\hat{r}=\sqrt{\hat{x}^2+\hat{y}^2}$. The general solution for S on or outside a control surface, which encloses all the sound sources under consideration, is given by

$$S(\hat{x},\hat{y}) = \sum_{n=0}^{\infty}(a_n\cos n\hat{\theta} + b_n\sin n\hat{\theta})H_n(k\hat{r}\beta^{-1}), \tag{6}$$

where H_n is the n-th-order Hankel function of the second kind. Combining the above equations yields

$$P(x,y) = \exp(ikM\hat{x}\beta^{-1})\sum_{n=0}^{\infty}(a_n\cos n\hat{\theta} + b_n\sin n\hat{\theta})H_n(k\hat{r}\beta^{-1}). \tag{7}$$

This is the general solution of 2D acoustic radiation with a uniform flow in x-direction. Note that $\hat{x}$ is scaled by the factor $\beta=\sqrt{1-M_a{}^2}$, $\hat{r}$ is not the usual $\sqrt{x^2+y^2}$, and $\hat{\theta}$ is also different from the usual angle θ in the polar coordinates for xy-plane.

In order to obtain the solution of Eq. (7), it is necessary to determine the coefficients a_n and b_n. These coefficients are determined by matching the assumed form of the solution to the measured acoustic pressure and its normal derivative over an open control surface. Once these coefficients are determined, the solution can be quickly evaluated at any field point on or outside the control surface. In the AIBM, both the acoustic pressures and its simultaneous, co-located derivative (along out normal direction) on the boundary Γ_1 are given as the input for the reconstruction of the acoustic field in the domain Ω (see Fig. 1). With the pressure derivative boundary condition as an additional input, the uniqueness of the reconstructed solution is guaranteed from the unique continuation

theory of elliptical equations. The method also yields a consistent and accurate solution on and outside of the control surface. When using the AIBM, it is assumed that the control surface is known although the exact locations of sound sources may not be available.

With the consideration of sound propagations in a uniform flow, the partial boundary value problem is defined as

$$\nabla^2 P - M_a{}^2 P_{xx} - 2ikM_a P_x + k^2 P = 0 \quad \text{in } \Omega = R^2 \setminus \Omega_{in},$$

$$P|_{\Gamma_1} = P, \quad \partial_{\mathbf{n}} P|_{\Gamma_1} = P\mathbf{n}, \tag{8}$$

where $\mathbf{n}$ is the outward normal to Γ_1. Similar to the procedures given in our earlier work,[4] the following steps are used to solve Eq. (8) in the solution form given by Eq. (7).

(a) Step 1: Finite summation approximation. The infinite summation in Eq. (7) is replaced by a finite summation, i.e.,

$$P(x,y) \approx \exp(ikM\hat{x}/\beta)\left[a_0 H_0(k\hat{r}/\beta) + \sum_{n=1}^{N} (a_n \cos n\hat{\theta} + b_n \sin n\hat{\theta}) H_n(k\hat{r}/\beta) \right], \tag{9}$$

where N is a suitable integer. The choice for N will be discussed later in this section. One obvious restriction is that the number of coefficients $(2N+1)$ to be determined must be less than the number of measurement points.

(b) Step 2: Hankel function evaluation. As there is no exact expression for Hankel function calculation, the recurrence formulations with the asymptotic H_0 and H_1 are used. They are applicable for the accurate calculation of Hankel function when $k\hat{r}\beta^{-1}$ is relatively large.

It is known that H_n has the following recurrence expressions:[9]

$$H_n(r) = \frac{2(n-1)}{r} H_{n-1}(r) - H_{n-2}(r), \quad n = 2,3,\ldots, \tag{10}$$

$$H_n'(r) = \tfrac{1}{2}[H_{n-1}(r) - H_{n+1}(r)], \quad n = 1,2,3,\ldots, \tag{11}$$

$$H_0' = -H_1. \tag{12}$$

In the current study, the H_0 and H_1 are calculated by the following asymptotic expansions with the first eight terms:

$$H_0(r) = \sqrt{\frac{2}{\pi r}} \exp(r - \pi/4) \sum_{j=0}^{7} (c_j r^{-j}), \tag{13}$$

$$H_1(r) = \sqrt{\frac{2}{\pi r}} \exp(r - 3\pi/4) \sum_{j=0}^{7} (d_j r^{-j}), \tag{14}$$

where the complex coefficients are

$$c_0 = 1, \quad c_j = \frac{-i(2j-1)^2}{8j} c_{j-1},$$

$$d_0 = 1, \quad d_j = \frac{-i(2j-3)(2j+1)}{8j} d_{j-1}. \tag{15}$$

(c) Step 3: Suitable coefficients optimization method. The partial boundary value problem, Eq. (8), is solved by the least squares technique. If P and $P_{\mathbf{n}}$ are known at M discrete points, $(r_1,\theta_1),\ldots,(r_M,\theta_M)$, the linear system for the coefficients a_n and b_n is given by the following $2M$ equations:

$$\exp(ikM_a\hat{x}_j\beta^{-1})\left(a_0 H_0(k\hat{r}_j\beta^{-1}) + \sum_{n=1}^{N} (a_n \cos n\hat{\theta}_j + b_n \sin a\hat{\theta}_j) H_n(k\hat{r}_j\beta^{-1}) \right) = P(x_j,y_j), \tag{16}$$

$$\partial_{\mathbf{n}}\left[\exp(ikM_a\hat{x}_j\beta^{-1})\left(a_0 H_0(k\hat{r}_j\beta^{-1}) + \sum_{n=1}^{N} (a_n \cos n\hat{\theta}_j + b_n \sin n\hat{\theta}_j) H_n(k\hat{r}_j\beta^{-1}) \right) \right] = P_{\mathbf{n}}(x_j,y_j), \tag{17}$$

where $j=1,2,\ldots,M$. It is noted that the terms $\partial\hat{r}/\partial\mathbf{n}$ and $\partial\hat{\theta}/\partial\mathbf{n}$ need to be evaluated first to solve above linear system.

(d) Step 4: Solving linear matrix equations. Expressing the linear system, Eqs. (16) and (17), in a matrix form as $AX=B$, where

$$X = [a_0,\ldots,a_N,b_1,\ldots,b_N]^T, \tag{18}$$

$$B = [P(x_1,y_1),\ldots,P(x_M,y_M),P_{\mathbf{n}}(x_1,y_1),\ldots,P_{\mathbf{n}}(x_M,y_M)]^T, \tag{19}$$

and $A=(A_1,A_2)^T$, A_1 corresponds to P and A_2 corresponds to normal derivative $P_{\mathbf{n}}$. One could find X by minimizing $\|AX-B\|^2$, where $\|\cdot\|$ is the standard L^2 norm. During the numerical study, it is observed that although some regularization methods may be needed to improve the stability of the system for large N, they are not necessary for relatively small N. In the current work, N is initialized within a given range of 1–30. The reconstructed solutions for various N are compared with the input data. The total error at all the input points is then computed for each N. The optimum N within the given range is determined from the minimum overall error. The reconstruction of the entire acoustic field is carried out using the optimum N.

It is worth mentioning that without the inclusion of the pressure gradient to the input, that is, only $P|_{\Gamma_1}$ specified, the solution of this partial boundary value problem is not unique. For example, if Ω_{in} is the unit disk centered at the origin, we assume that all the acoustic sources are enclosed by the unit circle, and the boundary Γ_1 is defined as $\Gamma_1 = \{(x,y)\,|\,\hat{x}^2+\hat{y}^2=\hat{r}^2, \hat{y}>0\}$. That is, Γ_1 is the upper half boundary of an ellipse, and $\hat{r}$ is chosen large enough so that the ellipse will include all the acoustic sources. The acoustic pressure for $\hat{\theta}\in[0,\pi]$ should satisfy Eq. (7). It is well-known that a_n and b_n are not unique from Fourier analysis. Especially P always has Fourier cosine expansion for $\hat{\theta}\in[0,\pi]$, namely, one can set $b_1=b_2=\cdots=0$. In any case, it is impossible to predict solutions for $\hat{\theta}\in(\pi,2\pi)$.

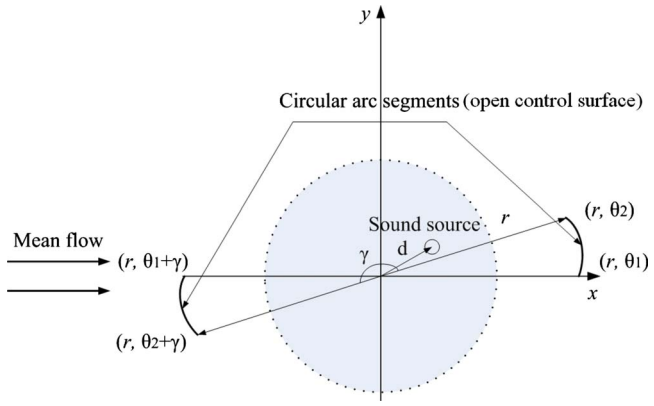


FIG. 2. (Color online) Schematic diagram of a monopole radiation in a uniform flow and locations of acoustic measurements.

Furthermore, it should be pointed out that even though our formulas are derived for the uniform flow $\mathbf{U}=U\hat{i}=(U,0)$, it is very easy to expand it for a general case $\mathbf{U}=(U,V)$. In that situation, the pressure $P(x,y)$ in the frequency domain with angular frequency ω can be achieved by rotating the coordinates for an angle α, where α in the range of $(0,2\pi)$ is given by $(U,V)=\sqrt{U^2+V^2}(\cos\alpha,\sin\alpha)$. Therefore, we have

$$P(x,y)=\exp(ikM_a\hat{x}\beta^{-2})\sum_{n=0}^{\infty}(a_n\cos n\hat{\theta}+b_n\cos n\hat{\theta})H_n(k\hat{r}/\beta), \tag{20}$$

where

$$M_a=\sqrt{U^2+V^2}/c,$$

$$\hat{x}=(x\cos\alpha+y\sin\alpha)/\beta=\hat{r}\cos\hat{\theta},$$

$$\hat{y}=-x\sin\alpha+y\cos\alpha=r\sin\hat{\theta},$$

$$\hat{r}=\sqrt{\hat{x}^2+\hat{y}^2}. \tag{21}$$

The AIBM is a frequency domain method. To deal with broadband noise propagation problems, the fast Fourier transform (FFT) needs to be used. The acoustic pressure for each frequency component can be individually calculated. A superposition of the contribution from each frequency gives the total acoustic pressure. The details of the AIBM implementation for broadband noise are explained in the study of sound scattering problem in Sec. III.

III. NUMERICAL EXAMPLES AND DISCUSSIONS

A. Monopole in a uniform flow

The monopole radiation in a subsonic flow is first studied. The monopole is placed at $(1\text{ m}, 0.2\pi)$ in polar coordinates. The wavenumber and intensity of the monopole are given as $k=2\text{ m}^{-1}$ and $A_0=0.001\text{ m}^2/\text{s}$. The uniform flow is in the $+x$-direction with $M_a=0.5$. A schematic diagram is shown in Fig. 2. Two circular arc segments are the locations of the acoustic input and the angle γ is a measure of the dimensionless distance between the two segments, with $\gamma=\pi$ being the farthest, when the segments are at the opposite sides of the origin. The units used for r and θ are meters and radians, respectively. The control surface considered is the circle of radius $r=2$ m.

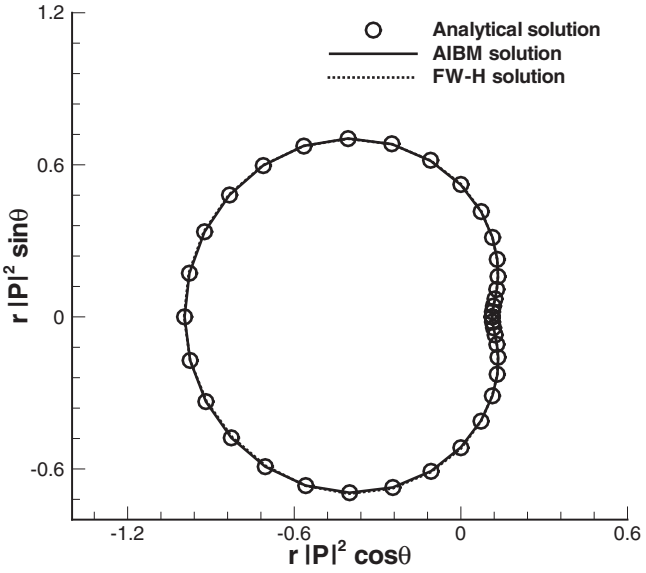


FIG. 3. Far-field directivity ($r=100$ m) comparison of a monopole radiation in a $M_a=0.5$ flow.

The analytical solution is used to provide the acoustic input data and verification for the AIBM. It is given by Dowling and Ffowcs Williams[10] in the form of a complex velocity potential as

$$\Phi(x,y,t)=A_0\frac{i}{4\beta}\exp[i(\omega t+M_akx\beta^{-2})]H_0\left(\frac{k}{\beta}\sqrt{\frac{x^2}{\beta^2}+y^2}\right). \tag{22}$$

For potential flow, the analytical acoustic pressure can be expressed as

$$p(x,y,t)=-\rho_0\left(\frac{\partial\Phi}{\partial t}+U\frac{\partial\Phi}{\partial x}\right). \tag{23}$$

Performing FFT, the acoustic pressure $P(x,y,\omega)$ in the frequency can be obtained. Its normal derivative $P_{\mathbf{n}}$, over the two circular arc segments, can be then derived. In this study, the polar coordinates of the starting and ending points of the two segments are (10 m, 0), $(10\text{ m}, 0.1\pi)$ and $(10\text{ m}, \gamma)$, $(10\text{ m}, 0.1\pi+\gamma)$, respectively. Ten uniformly spaced grid points are used on each of the segments. With $\gamma=\pi$, the far-field directivity from the AIBM is calculated and compared with those from the analytical solution and the FW-H integral equation in Fig. 3. The input acoustic data on a closed surface with $r=10$ m are used for the FW-H method. The results show excellent agreement among these three methods. It is important to point out that the arc length of the each input segment for the AIBM is only about 1/20 of the circumference of the FW-H surface.

As it has been shown in the work of Yu *et al.*,[4] though an accurate reconstruction can be obtained from the input given over an open surface, the AIBM becomes less effective when the input segments become clustered. As a general rule, the more scattered the input segments around the sound sources,

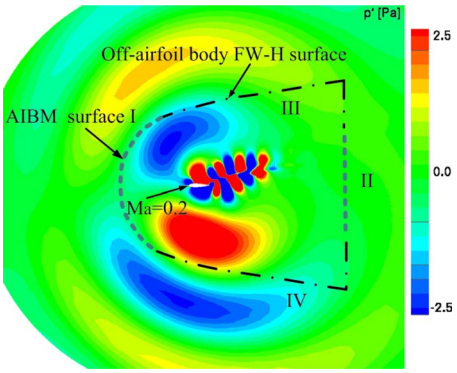


FIG. 4. (Color online) Instantaneous pressure perturbations of the flow around NACA 0018 airfoil along with the location of the FW-H surface.

the more accurate the reconstructed acoustic solution. Since the choices of the input segments are limited by the accessibility and practicality of the acoustic measurement in the radiated field, some regularization techniques may be needed to improve the effectiveness of the AIBM when the input locations are not scattered far enough around the sound sources.

B. Aerodynamic sound radiation of a flow around a NACA airfoil

After the successful verification of the AIBM, the method is coupled with CFD simulation to predict the sound radiation by a uniform flow around a NACA airfoil. The instantaneous acoustic pressure contour from CFD simulation is shown in Fig. 4 for NACA0018 airfoil of a chord length 0.3 m. The free-stream Mach number, M_a, is 0.2 and the angle of the attack is 20°. The details of the CFD solutions were given by Greschner *et al.*[11] In this study, both the AIBM and FW-H methods will be used and compared for the far-field acoustic prediction.

The far-field acoustic solution is commonly obtained by solving the FW-H integral equation based on the unsteady CFD data from a FW-H surface that completely encloses the airfoil. The FW-H equation is a rearrangement of the exact continuity and momentum equations to a wave equation with source contributions from the monopole, dipole, and quadrupole terms. The contribution of the quadrupole term, the Lighthill stress tensor, is usually neglected since the FW-H surface, as indicated in Fig. 5 for the current study, is placed outside of the region where the stress tensor is significant.

The AIBM is carried out by using the unsteady pressure and pressure derivative data over an open surface, formed by the curved segments I and II or segments III and IV of the FW-H surface (see Fig. 4), as the acoustic input. The two-dimensional formulation of the FW-H equation in the frequency-domain[12,13] is used with the input of the unsteady pressure and velocity solutions over the entire FW-H surface.

The comparison of the far-field directivity obtained from the AIBM and the FW-H method is shown in Fig. 5. As can be seen in Fig. 5, the results of the AIBM based on the inputs of the two chosen open surfaces agree reasonably well with that of the FW-H method. In order to have an overview of the acoustic propagation in the far field, the sound pressure

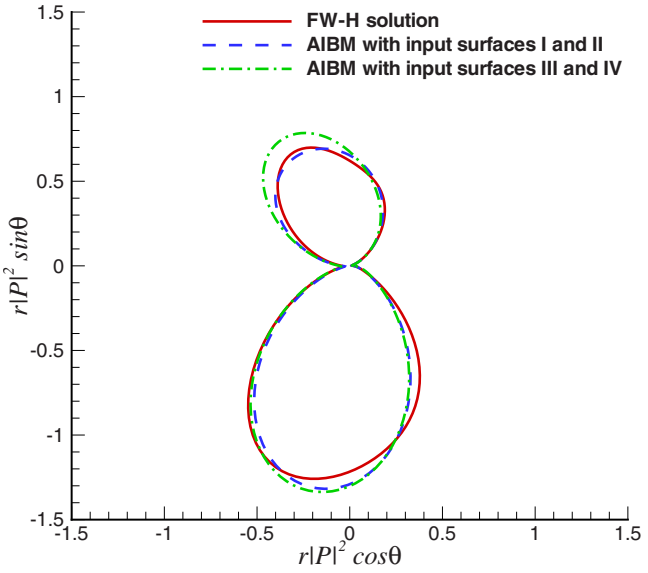


FIG. 5. (Color online) Far-field directivity (r=20 m) of the flow around NACA 0018 airfoil.

contour plots from the AIBM and the FW-H method are shown in Fig. 6. It is noted that the radius of the control surface shown in the figure is r=3 m, which is ten times of the chord of the airfoil. It should also be pointed out that the FW-H surface used in the study, though not a circular shape, is enclosed in this control surface. The close agreement among the contour plots indicates that the AIBM is capable to effectively obtain the radiated acoustic field based on the acoustic input from an open control surface. The method, therefore, has a potential application for the far-field acoustic reconstruction for problems where a closed FW-H surface is not possible.

C. Sound scattering by a circular cylinder

This example is an ideal model of the physical problem of predicting the sound field generated by a propeller scattered off by the fuselage of a moving aircraft.[14] In the model, the fuselage is considered as a circular cylinder and the noise source (propeller) as a line source such that the computational problem is two-dimensional. A polar coordinate system and the Cartesian coordinate system centered at the center of the circular cylinder of dimensionless radius 0.5 (the diameter of the cylinder of 1 m is used as the reference length) are shown in Fig. 7. The mean flow is given as Mach number of zero. The governing equations for this problem are the linearized dimensionless Euler equations. All variables considered here are dimensionless and the speed of sound is used as the reference velocity. The equations are discretized using the optimized upwind dispersion-relation-preserving scheme of Zhuang and Chen.[15] The detailed implementations of the boundary and initial conditions are given by Chen and Zhuang,[16] in which the numerical solution (CAA solution) was also verified by the analytical solution. In the current study, the CAA solution is used as the acoustic input for the AIBM. The input is given at 40 uniformly distributed points on each of the two circular seg-

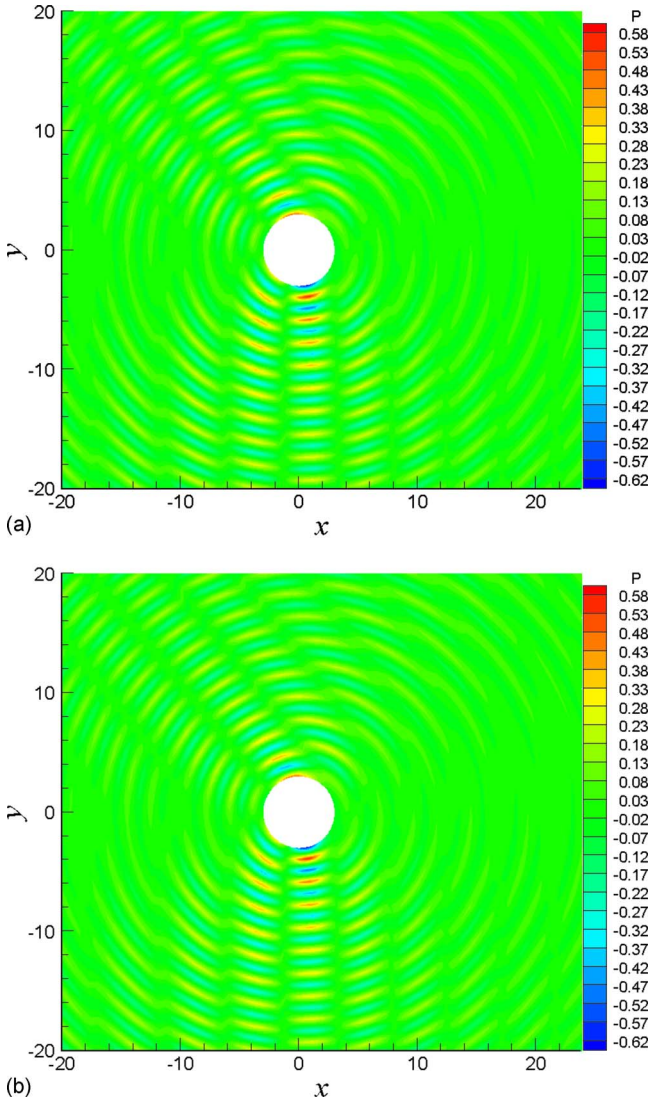


FIG. 6. (Color online) Pressure contours of the sound propagation generated by the flow around NACA 0018 airfoil: (a) FW-H and (b) AIBM.

ments (see Fig. 7) with a dimensionless radius of 6.125. The initial pressure pulse located at (4,0) is given as

$$p(x,y) = \exp\left[-\ln 2\frac{(x-4)^2+y^2}{0.2^2}\right] \tag{24}$$

and the perturbation velocity components in x- and y-directions are considered as zero, $u=v=0$.

The input acoustic data are obtained from numerical solutions of CAA on a circle with radius r_0 for $0 \le t \le t_0$, where t_0 is large enough so that the solution asymptotically decays at large t. Assuming the dimensionless speed of sound $c=1$, then the wave number for each frequency $k_j=\omega_j$. The FFT is then used to decompose the solution in terms of its frequencies ω_j, i.e.,

$$p(x,y,t) = \sum_{j=1}^{J} \exp(i\omega_j t) P_j(x,y), \tag{25}$$

with $P_j(x,y)$ satisfying the Helmholtz equation

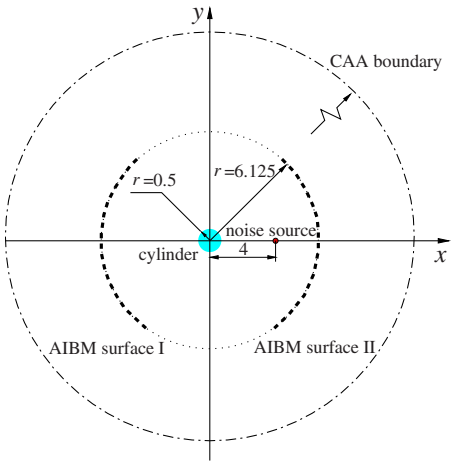


FIG. 7. (Color online) Schematic diagram of sound scattering by a cylinder.

$$\nabla^2 P_j + \omega_j^2 P_j = 0, \quad x \in \Omega = R^2 \setminus \Omega_{in}. \tag{26}$$

If the polar coordinates are used in a two-dimensional configuration, the solution of Eq. (26), which also satisfies the radiation condition, can be written as

$$P_j(x,y) = \sum_{t=0}^{\infty} (a_{j,n}\cos n\theta + b_{j,n}\sin n\theta) H_n(\omega_j r). \tag{27}$$

The numerical solution of Eq. (27) is obtained by replacing infinite summation in Eq. (27) by a finite summation, i.e.,

$$P(x,y) \approx a_{j,0}H_0(\omega_j r) + \sum_{n=1}^{N_j} (a_{j,n}\cos n\theta + b_{j,n}\sin n\theta) H_n(\omega_j r). \tag{28}$$

The acoustic pressure in the time domain is then determined by the superposition of each frequency

$$p(x,y,t) \approx \sum_{j=1}^{J}\sum_{n=0}^{N_j} (a_{j,n}\cos n\theta + b_{j,n}\sin n\theta) e^{i\omega_j t} H_n(\omega_j r), \tag{29}$$

where N_j is a suitable integer for each frequency component. As in the case for single frequency, these coefficients are determined by matching the assumed form of the solution to the input acoustic data in the frequency domain. The N_j must be carefully chosen to ensure both the accuracy and stability as discussed in Ref. 4. Apparently, the accuracy of the solution naturally requires a large value of N_j. However, a large N_j can also result in computational instability. For any given ω_j, a mathematical analysis yields

$$H_n(\omega_j r) = \left(\frac{2n}{e\omega_j r}\right)^n, \quad n \to \infty. \tag{30}$$

When $\omega_j r$ is small, $H_n(\omega_j r)$ is very large when $n \approx (e\omega_j r/2)$. A small perturbation in the input acoustic data may result in large variations in $a_{j,n}$ and $b_{j,n}$. Hence the value for N_j has to be restricted for the stability of the numerical solution (Fig. 8). In the current work, the value of N_j is chosen as

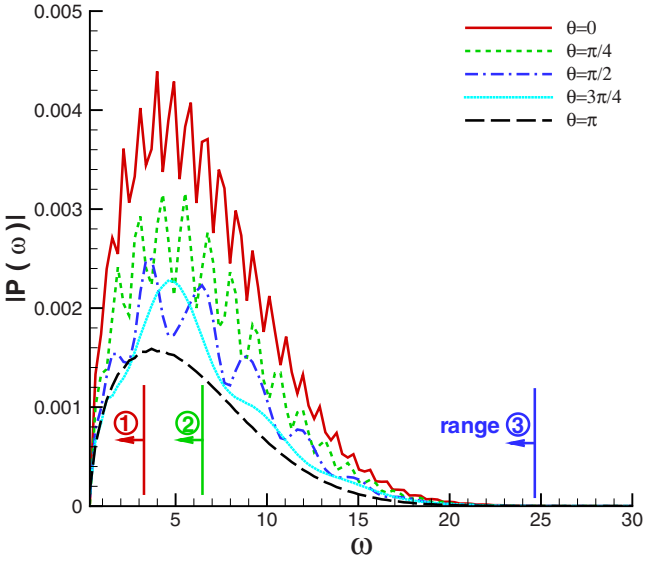


FIG. 8. (Color online) Frequency spectrum for sound scattering, $r=6.125$.

$$N_j = \min\{[\omega_j r_0] + 1, 30\} \tag{31}$$

with considerations of numerical stability and computational efficiency. [ ] in the above equation represents the greatest integer less than or equal to the term inside. For any given ω_j and N_j, the coefficients $a_{j,n}$ and $b_{j,n}$ are determined for each frequency component of $1 \le j \le J$ and $0 \le n \le N_j$ based on the acoustic input provided on the circle of r_0.

The results of predicted sound pressure time history for different frequency ranges are also compared with the corresponding CAA solutions at $x=0$ in Fig. 9. As it is indicated the accuracy of the predicted solution improves significantly as all the dominant frequencies are included in the AIBM calculation. In terms of peak pressure values and locations of the incident and reflected waves, excellent agreements between the two methods, AIBM and CAA, are demonstrated. The oscillations at the lower amplitudes are due to a relatively large value of N. As it is discussed in Sec. II, the number of summations, N in Eq. (8), needs to be reduced as the radius r decreases to provide a converged solution. The instantaneous pressure contour plots from the CAA calculation and the AIBM are compared in Fig. 10. The results of these contour plots demonstrate that the AIBM can effectively predict the propagations of both the incident and reflected sound waves.

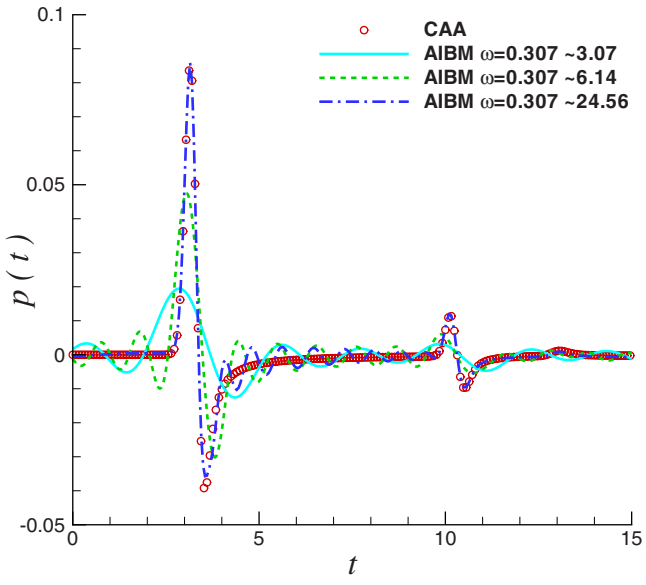


FIG. 9. (Color online) Pressure time history reconstruction with different frequency range at $r=7.25$ and $\theta=0°$.

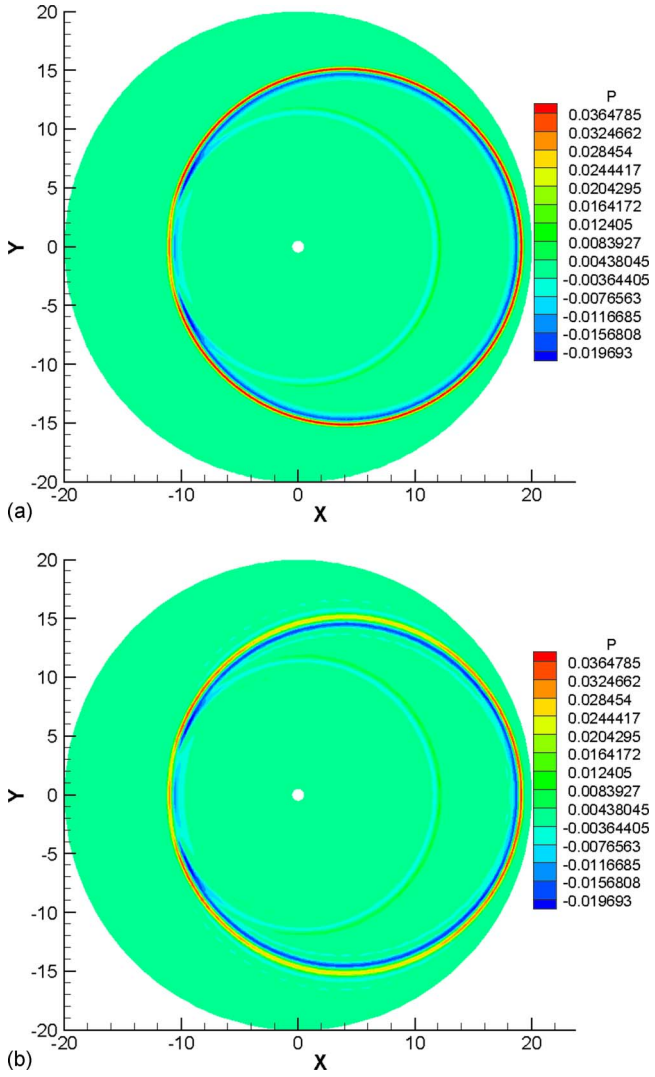


FIG. 10. (Color online) Instantaneous pressure contours of sound scattering at $t=15$: (a) CAA and (b) AIBM.

IV. CONCLUSIONS

An AIBM is successfully formulated, implemented, and verified for predicting sound propagations in uniform flows. The results of flow around the airfoil show that the AIBM can predict the far-field acoustic propagations accurately based on the acoustic input over an open control surface. The method is therefore particularly useful for aeroacoustic applications, where the far-field acoustic measurement over a closed surface enclosing all the sound sources under consideration is infeasible. It needs to be pointed out that the open surface, where the input is given, should be far enough from the aerodynamic sources so that from thereon to the far field

only the sound waves propagate in a uniform flow. Furthermore, the accurate prediction of the sound propagation and reflection of the scattering problem demonstrates that the AIBM can be effectively used for broadband noise problems. Due to the efficiency advantage of the AIBM, the method has potential to become a part of an integrated computational procedure for prediction of far-field sound propagations of practical aeroacoustic applications. As pointed out by Ffowcs Williams (1993) that the nature of aeroacoustic fields "permits many different but equally exact computational procedures for evaluation both the sound and its source field." The current study demonstrates that the AIBM can be used for the far-field acoustic reconstruction for aeroacoustic problems when a closed FW-H control surface is not possible.

ACKNOWLEDGMENTS

This research is partially financial supported by the National Science Foundation under Grant No. NSF-ITR-0325760. The first author (C.Y.) would like to acknowledge the financial support from the Technical University of Berlin.

[1]J. E. Ffowcs Williams and D. L. Hawkings, "Sound generated by turbulence and surfaces in arbitrary motion," Philos. Trans. R. Soc. London, Ser. A **264**, 321–342 (1969).
[2]F. Farassat and M. K. Myers, "Extension of Kirchhoff's formula to radiation from moving surfaces," J. Sound Vib. **123**, 451–461 (1988).
[3]P. Di Francescatonio, "A new boundary integral formulation for prediction of sound radiation," J. Sound Vib. **202**, 491–509 (1997).
[4]C. Yu, Z. Zhou, and M. Zhuang, "An acoustic intensity-based inverse method for reconstruction of radiated fields," J. Acoust. Soc. Am. **123**, 1892–1901 (2008).
[5]Z. Wang and S. F. Wu, "Helmholtz equation least-squares (HELS) method for reconstructing the acoustic pressure field," J. Acoust. Soc. Am. **102**, 2020–2032 (1997).
[6]C. Yu, Z. Zhou, M. Zhuang, X. D. Li, and F. Thiele, "Effective far-field acoustic prediction method and its computational aeroacoustics applications," AIAA J. **47**, 410–417 (2009).
[7]C. Yu, Z. Zhou, and M. Zhuang, "Improved inverse acoustic methods through advances in acoustic intensity measurement techniques," AIAA Paper No. 2007-3563 (2007).
[8]C. Yu, Z. Zhou, M. Zhuang, X. D. Li, and F. Thiele, "3D acoustic intensity-based method and its CAA applications," AIAA Paper No. 2008-3053 (2008).
[9]F. Bowman, *Introduction to Bessel Functions* (Dover, New York, 1958).
[10]A. P. Dowling and J. E. Ffowcs Williams, *Sound and Sources of Sound* (Ellis Horwood, Chichester, Sussex, 1983).
[11]B. Greschner, C. Yu, S. Zheng, M. Zhuang, Z. J. Wang, and F. Thiele, "Knowledge based airfoil aerodynamic and aeroacoustic design," AIAA Paper No. 2005-2968 (2005).
[12]D. P. Lockard, "A comparison of Ffowcs Williams-Hawkin solvers for airframe noise applications," AIAA Paper No. 2002-2580 (2002).
[13]C. Yu and X. D. Li, "Far sound field prediction of a 2-D parallel shear layer basing on an integration methodology," J. Engineering Thermophysics (in Chinese) **24**, 939–942 (2003).
[14]Second Computational Aeroacoustics (CAA) Workshop on Benchmark Problems, Tallahassee, FL (1996), NASA CP-3352.
[15]M. Zhuang and R. Chen, "Applications of high-order optimized upwind schemes for computational aeroacoustics," AIAA J. **40**, 443–449 (2002).
[16]R. Chen and M. Zhuang, "Application of dispersion-relation-preserving scheme to the computation of acoustic scattering in benchmark problems," in ICASE/LaRC Second Workshop on Benchmark Problems in Computational Aeroacoustics, Tallahassee, FL, (1996).

Modeling of acoustic penetration into sandy sediments: Physical and geometrical aspects

V. Aleshin[a] and L. Guillon
Institut de Recherche de l'Ecole Navale (IRENav), BCRM Brest, CC 600, F-29240 Brest Cedex 9, France and Ecole Nationale Supérieure des Ingénieurs des Etudes et Techniques d'Armement (ENSIETA), E^3I^2-EA3876, 2 rue François Verny, F-29806 Brest Cedex 9, France

(Received 28 January 2009; revised 1 September 2009; accepted 1 September 2009)

Two different approaches to the problem of acoustic penetration into sandy marine sediments are considered: application of the Buckingham constitutive model for sediment with a plane surface and boundary element analysis of a rough surface of sediment represented as a homogeneous fluid. By a careful modeling of the constitutive behavior for plane seafloors, it is possible to partly reproduce some features of known experimental dependencies for acoustical pressure. However, accounting for roughness appears to be more important. Accordingly, the authors present a detailed numerical analysis of penetration into rough sediments using the boundary element method. The simulation results support conclusions reached by other investigators and demonstrate how local surface irregularities violate the evanescence condition that holds for a plane interface at subcritical incidence, thus considerably increasing penetration. The results apply to the frequency range 0.5–50 kHz and grazing angles larger than approximately 6°–8° at 10–50 kHz. For lower frequencies, when diffraction becomes important, the lowest possible grazing angle strongly depends on the range covered by the incident beam and is, in general, considerably larger. The authors provide several characteristic examples with frequencies 5 and 15 kHz and grazing angles 15°–30° illustrating the impact of roughness on penetration.
 [DOI: 10.1121/1.3238255]

PACS number(s): 43.30.Hw, 43.30.Ma [RCG] Pages: 2206–2214

I. INTRODUCTION

Models for penetration of sound into sandy marine sediments are typically of interest for detection of buried objects (pipelines, mines, archeological finds). This problem has two essential aspects.

- *Physical.* Physical properties of marine sediments, appropriate constitutive models, stratification problem (the sediment properties can have unknown depth dependencies).
- *Geometrical.* Bottom roughness (here understood as a geometrical aspect) and the associated acoustical scattering, sound beam directivity, and diffraction; accounting for small grazing angles at which, for instance, geometrical shadows at the surface can appear, etc.

From a physical point of view, marine sediment can be regarded as a viscous fluid or solid, as a saturated porous material, or as a granular material with internal friction, saturated by fluid. The first is the simplest case traditionally considered in underwater acoustics.[1] The second point of view arose as a natural extrapolation of the Biot theory onto saturated granular materials[2] and provides a more precise estimation of sound speeds at low frequencies hardly accessible for direct measurements. Its disadvantage is in the high number of parameters; most of them concern a hard skeleton or frame of a porous material, whose existence is doubtful in the case of unconsolidated granulars. In addition, the second "slow" Biot wave predicted by the theory has never been detected with full confidence[2–4] and, even if exists, must represent a minor effect anyway.

The latter (third) approach is based on the assumption that, for marine sediments, intergranular friction is more important than fluid viscosity. The corresponding grain-shearing (GS) model by Buckingham[5,6] enables one to calculate longitudinal and shear sound velocities and attenuations as functions of frequency and depth starting from only one important parameter: the grain size. Depth dependence in that case appears due to gravity: Grains are compressed by their own weight the same way as for dry granular materials.[7–11] Another advantage of the approach is a possibility to estimate sound velocities at low frequencies (1–20 kHz) using a dispersion law commonly found for materials with internal friction: attenuation linear with frequency[12] and sound velocity having logarithmic frequency dependence,[13] in accordance to the Kramers–Kronig relationships.

This recent approach provides an opportunity to reconsider known experimental results[14,15] on acoustic penetration and see whether use of a deeper theoretical description for the nature of the sediment better matches the experiments. Such an attempt is proposed in Sec. II of the paper, where we present our simulation for penetration of plane waves into sediments with a plane interface using the GS model. The results do not differ much from the classical case of fluid sediment, except in vicinity of the critical regime where strong gradient-induced effects are found.

[a]Author to whom correspondence should be addressed. Electronic mail: aleshinv@mail.ru

Accounting for geometrical factors of the problem, such as interface roughness, appears to be more important. If the wavelength is smaller than the size of surface irregularities, then locally, at the scale of a ripple, the character of penetration resembles the plane-wave case, but with the local grazing angle higher or lower than the nominal grazing angle. In that way, for some sections of the roughness profile, the local grazing angle can be higher than critical; the presence of such sections greatly enhances penetration. The idea about the extremely high impact of roughness was expressed by other authors;[14–17] our research supports their conclusion by virtually exact calculations in the subcritical regime. A detailed review of different factors contributing into the anomalous sound penetration for subcritical angles can be found in Ref. 18 and references therein.

Methods accounting for interface roughness include perturbation approach (analytical[1] or numerical[19] solutions), exact solutions requiring numerical implementation,[16] and asymptotical expansions such as the Helmholtz–Kirchhoff approximation[20] also known as the method of tangent plane.[1] In order to correctly represent the case of small grazing angles in which geometrical shadows can appear at the surface, especially behind high and sharp ripples, the exact methods are most adequate. Such a procedure based on integral equations with Green's functions has been proposed[21] for optical reflection from a random grating and then used[16] in the acoustical case. In fact, the same principle derived from the second Green's identity underlies most of commercial codes implementing the boundary element method (BEM).[22] In Sec. III we apply this method for homogeneous fluid sediment with parameters calculated from the GS model removing its shear component and flattening the depth dependencies. The chosen two-dimensional (2D) representation of the method makes it possible to evaluate the principal effects occurring due to roughness, keeping, at the same time, the computational expenses at a moderate level. As a result, we get virtually exact acoustical pressure fields corresponding to particular realizations of roughness and to a set of the incident beam parameters. The problem of geometrical shadows does not arise in that case, since the correct boundary conditions at the interface are used. The BEM code has been run with various combinations of physical and geometrical parameters of the system that enabled us to analyze different penetration regimes, including those in which the presence of roughness provides a huge gain in penetration. We compare the simulated penetration ratio to experimental results[14,15] and, at the end of Sec. III, present statistical interpretation of the results. The algorithm and the results presented here can help establish an optimal regime for real underwater systems designed for detection of buried objects.

II. PHYSICAL ASPECTS AND 1D PROBLEM

A. Buckingham's constitutive model

In the first part of our study, the complex P-wave and shear moduli M and G of sediment are given by the Buckingham GS model[5,6]

$$M = \rho c_0^2 + (\gamma_L + (4/3)\gamma_S)(i\omega T)^n, \quad G = \gamma_S (i\omega T)^n, \tag{1}$$

where ω is the angular wave frequency. All parameters in Eq. (1) can be calculated from material constants, scaling constants fitted for a large number of sediments, grain size a_g, and depth y in the sediment. In particular, density ρ of sediment may be expressed in terms of the porosity N as a weighted mean of the pore water density ρ_w and the density of mineral grains ρ_g as follows:

$$\rho = N\rho_w + (1-N)\rho_g. \tag{2}$$

The sound velocity c_0 in the absence of frictional effects, i.e., when $\gamma_L = \gamma_S = 0$ in Eq. (1), is expressed as $c_0 = \sqrt{K_0/\rho}$ through the bulk modulus of sediment K_0. The latter can be found from Wood's equation similarly to Eq. (2) as follows:

$$\frac{1}{K_0} = N\frac{1}{K_w} + (1-N)\frac{1}{K_g}, \tag{3}$$

with K_g, bulk modulus of grains, and K_w, bulk modulus of water.

Buckingham[5] proposed the compressional and shear coefficients $\gamma_{L,S}$ in Eq. (1) to be scaled with the grain size a_g, porosity, and depth y as

$$\gamma_L = \gamma_{L0}\left[\frac{(1-N)a_g y}{(1-N_0)a_0 y_0}\right]^{1/3}, \quad \gamma_S = \gamma_{S0}\left[\frac{(1-N)a_g y}{(1-N_0)a_0 y_0}\right]^{2/3}, \tag{4}$$

in which N_0, a_0, and y_0 are constants. Such dependencies arise due to the fact that grain interactions consist of a multitude of individual micro-slip events between asperities, so that $\gamma_{L,S}$ depend on the rates at which sliding events happen within a zone of contact between two grains.[6] The number of sliding events is proportional to the contact's perimeter for the normal compression and to the contact area for shearing. The contact radius is given by the Hertz solution as a function of the compression force that eventually provides the expected scaling law (4).

Finally, porosity N of the sediment can be obtained as a function of the grain size a_g assuming that the sediment is a random packing of identical spheres with rms roughness Δ, the same for different a_g. In that case, each sphere has different external and mean radii; the external one contributes to the volume of the packing, while the mean one contributes to its mass. Such considerations yield

$$N = 1 - P_s\left(\frac{a_g + 2\Delta}{a_g + 4\Delta}\right)^3. \tag{5}$$

Here Δ is a value of order of 1 μm ($\Delta = 1$ μm frequently used[5]) and $P_s \approx 0.63$ is a constant of random packing. In fact, its value can differ since it is affected by at least two factors. On one hand, size variations of particles around the mean magnitude a_g usually break order of packing thus reducing P_s and increasing porosity N. On the other hand, very small particles (clay and silt) always present in the mixture can fill in the pores between larger particles and effectively increase P_s.

Other constants ($\gamma_{L0} = 3.888 \times 10^8$ Pa, $\gamma_{S0} = 4.588 \times 10^7$ Pa, $n = 0.0851$, $T = 1$ s, $a_0 = 1$ mm, $y_0 = 0.3$ m, and $N_0 = 0.377$) were fixed[5] to match an extensive series of experimental results.

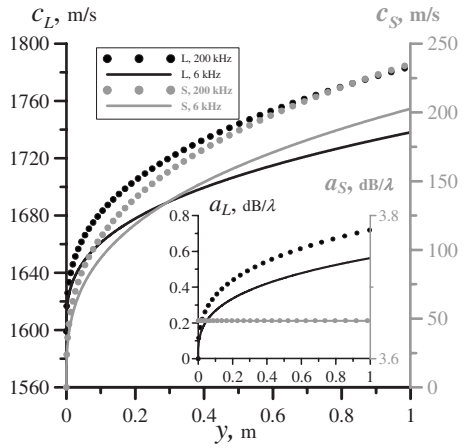


FIG. 1. Buckingham GS model: depth profiles of longitudinal (left axis, in black) and shear (right axis, in gray) velocities at 6 (solid lines) and 200 kHz (dots). The choice of frequencies corresponds to the experiments (Ref. 14). The inset shows attenuations for longitudinal and shear waves in the same way.

In our simulation, we used physical parameters $K_w=2.34$ GPa, $K_g=36$ GPa, $\rho_w=1024$ kg/m^3, and $a_g=0.2$ mm relevant to the experimental situation.[14] Other data available for this type of sediment are porosity 44.6%, density 1920 kg/m^3, and average longitudinal sound speed 1720 m/s measured at 200 kHz and at 30 cm depth. These values can be matched simultaneously, for instance, by taking $\rho_g=2650$ kg/m^3 (a value close to $\rho_g=2670$ kg/m^3 documented[15] for the same type of sediment) and by choosing $P_s=0.571$ instead 0.63, as it was advised by Buckingham.[5] Such an adjustment produced the following result: The same porosity $N=44.6\%$ and sound velocity $c_L=1727$ m/s with the density $\rho=1925$ kg/m^3 slightly different from the measured values. More data on the sediment would reduce the ambiguity in choice of parameters; however, even this example provides a satisfactory match within a typical error level.

The depth profiles of the longitudinal and shear sound velocities $c_{L,S}$ and attenuations $a_{L,S}$ are plotted in Fig. 1 for two frequencies: 200 and 6 kHz. The shear attenuation a_S expressed in dB/λ is independent of frequency and depth. Dispersion produced by the GS model yields at 30 cm depth and frequency of 6 kHz the value of 1690 m/s for the longitudinal sound velocity, which is quite close to the estimation[14] of 1685 m/s based on the Biot theory. The power depth dependencies for the sound velocities resulting from the proposed scaling equation (4) are quite typical for marine sediments[23,24] or unconsolidated dry granular materials under gravity[7–10] where guided surface acoustic modes (GSAMs) (Refs. 8 and 9) were detected, but actual powers can vary.

By fitting the velocity profiles $c_{L,S}(y)$ from Fig. 1 by functions $c_L=c_{L0}+c_{L1}y^{\beta_L}$ and $c_S=c_{S1}y^{\beta_S}$, one obtains powers $\beta_S=1/3$ and $\beta_L\approx 0.35$. Our fits of velocity profiles measured by Hamilton[23] give $\beta_L\approx 0.26$. Measurements for shear velocities[24] provide values α_S ranging from 1/6 for very high pressures to 1/3 for weakly compressed sediments, with the most typical observations around 1/4. A recent theory[10] for dry granular materials provides the powers 1/3 for the bulk modulus and 2/3 for the shear modulus, i.e., exactly the same as the Buckingham GS model [Eq. (4)], while the classical study[11] of an ensemble of Hertz spheres indicates 2/3 for both moduli. Direct measurements of the velocities against pressure are most reliable for high pressures and frequently yield[7] powers about 1/4. Indirect measurements[9] of the velocity profiles via GSAMs, in which measured surface data were fitted by varying the parameters of these profiles, suggest powers $\beta_L=0.305$ and $\beta_S=0.32$ for glass beads of 150 μm diameter.

In that way, the agreement between the GS model and data established in several reference points enables us to hope that the model is globally adequate, and that the rigidity profiles close to those depicted in Fig. 1 take place in reality.

In Sec. II B, we write and numerically solve the Helmholtz equations that have the same form as for GSAMs, together with boundary conditions that are, however, different.

B. 1D problem of acoustic penetration into Buckingham's medium

In the one-dimensional (1D) geometry, i.e., for a plane wave incident on a plane interface, the acoustical displacement has the form $(u_x(y),u_y(y))\exp(i\omega t-i\kappa x)$ where the amplitudes u_x and u_y depend on depth y only, x is the horizontal coordinate, ω is the wave cyclic frequency, and κ is the wavenumber corresponding to the harmonic dependence on x. The Helmholtz equation[8,9,25] for this displacement reads

$$(G(y)u_x')' + (\rho(y)\omega^2 - k^2M(y))u_x = i\kappa[(M(y)-2G(y))u_y' + (G(y)u_y)'],$$

$$(M(y)u_y')' + (\rho(y)\omega^2 - k^2G(y))u_y = i\kappa[G(y)u_x' + ((M(y)-2G(y))u_x)'], \quad (6)$$

where moduli M and G are given by Eq. (1). Equation (6) must be supplemented with the appropriate boundary conditions corresponding to the absence of stress σ_{xy} at the surface, continuity of the stress σ_{yy} and the displacement u_y at the surface, and vanishing the solution at $y\to\infty$. The latter condition is assured at least due to dissipation present in the GS model.

The absence of shear stress σ_{xy} at the surface gives a simple condition

$$[G(y)(u_x' - i\kappa u_y)]|_{y=0} = 0. \quad (7)$$

Continuity of the compressive stress σ_{yy} and displacement u_y requires an explicit form for the incident wave

$$\vec{u}_i = u_0(\cos\alpha, \sin\alpha)\exp(i\omega t - ik_w x\cos\alpha - ik_w y\sin\alpha), \quad (8)$$

and for the reflected wave

$$\vec{u}_r = u_0R(\cos\alpha, -\sin\alpha) \times \exp(i\omega t - ik_w x\cos\alpha + ik_w y\sin\alpha), \quad (9)$$

with R the complex reflection coefficient, and $k_w=\omega/c_w$ the wavenumber in water. Then, continuity of σ_{yy} and u_y provides the condition $\kappa=k_w\cos\alpha$ together with

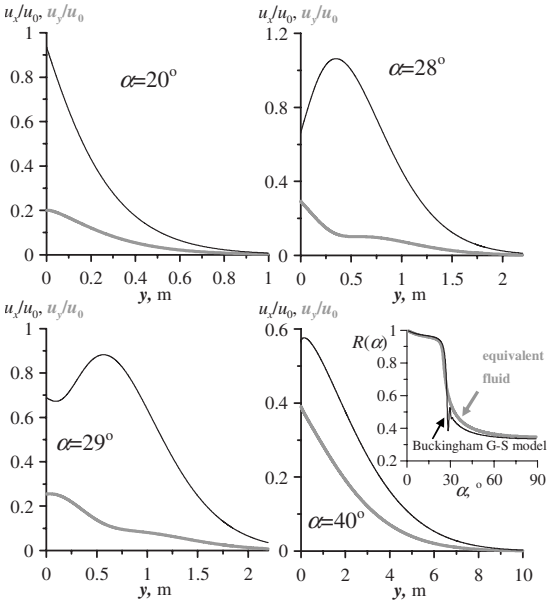


FIG. 2. Plane-wave penetration into Buckingham's medium: depth profiles of displacements u_x (black) and u_y (gray) for different grazing angles and $f=4$ kHz. Angles $\alpha=28°$ and $\alpha=29°$ are in vicinity of critical regime. The inset shows the angular dependence of the reflection coefficient $R(\alpha)$ for the Buckingham GS model and fluid sediment with parameters from experiments (Refs. 14 and 15). Critical regime corresponds to the range $\alpha=26°-32°$ when the propagative transmitted and evanescent waves are both generated and effectively interfere.

$$[M(y)u_y' - i\kappa(M(y) - 2G(y))u_x]|_{y=0} = -ik_w K_w u_0(1+R) \tag{10}$$

and

$$u_y(0) = u_0(1-R)\sin\alpha. \tag{11}$$

Finally, vanishing of acoustic waves at infinite depth means that

$$u_{x,y}|_{y\to\infty} = 0. \tag{12}$$

So, the system of Helmholtz equations (6) with coefficients in the form of Eq. (1), together with boundary conditions (7) and (10)–(12) describes completely the reflection and penetration processes.

C. Results for 1D penetration problem and discussion

A single differential equation of the second order corresponds, after the finite-difference discretization, to a linear algebraic system with a tridiagonal matrix. Such systems are easy to solve numerically with the help of the tridiagonal matrix algorithm (TDMA),[26] also known as the Thomas algorithm. Here, since we have Eq. (6), which is already a system of equations, a direct generalization of the TDMA was used where each element of the tridiagonal matrix is a 2×2 matrix itself.

The numerically calculated displacement profiles are plotted in Fig. 2 as functions of depth y: Four graphs show the u_x (black) and u_y (gray) displacements for different grazing angles. The critical regime is unambiguously defined for homogeneous sediment only, since only in that case the depth dependencies are given by simple exponential functions,[1] and the corresponding wavenumber—real or complex—is clearly defined. However, far from the critical regime, the displacement profiles are almost exponential, despite the inhomogeneity of the sediment: For subcritical angles ($\alpha=20°$ in Fig. 2) the exponential decay corresponds to the evanescent wave, while for supercritical angles ($\alpha=40°$ in Fig. 2) it is related to physical attenuation of the transmitted wave. The other two sets in Fig. 2 illustrate the critical regime that occurs approximately at $\alpha=26°-32°$. The behavior of the displacement profiles is then more complex: Oscillations in the vicinity of the surface show up, and the profiles become much more sensitive to the grazing angle α. These oscillations can be interpreted as interference of the evanescent and transmitted waves both existing in that case, since the rigidity at different depths is different.

The oscillations in the displacement profiles near the critical regime lead to rapid variations of the reflection coefficient R through Eq. (11). The angular dependence of R is given in the small inset in Fig. 2. In the present case of Buckingham's medium, the behavior of the reflection coefficient for small and large angles is close to that of an equivalent fluid, which is defined simply as a homogeneous material with constants c_L and a_L taken from the profiles in Fig. 1 at some characteristic depth (in our example in the inset in Fig. 2, $y=0.2$ m), and $c_S=a_S=0$. However, as it was mentioned, in the critical regime when the evanescent and transmitted waves are both present, the behavior of the reflection coefficient is largely different.

Comparison of the calculated penetration ratio (pressure amplitude in sediment referenced by the incident field at the surface) for the sediments with and without gradients (Fig. 3) to known experimental results[14] shows that the slopes of the corresponding experimental and theoretical curves generally coincide, but there are constant shifts frequently observed between them. Shifts between the theoretical curves are usually less than between theory and experiment, especially at 30 cm depth, since our constant velocity taken at 20 cm roughly represents an average for $0<y<30$ cm. The shifts between theoretical curves should be attributed mostly to gradients in the GS model. In Fig. 3 one can see that the fluid sediment model sometimes provides slightly better results than the GS model accounting for gradients. Later on we will see that even for low frequencies 2–6 kHz roughness-induced variability in data is very strong and effectively masks the mismatch between the theoretical curves. In that way, using the available data[14] and our match in Fig. 3, it is difficult to judge if the rigidity gradients in the GS model are realistic or not; however, the cited experimental observations and theories enable the hope that gradients provided by the GS model are realistic in many cases. More precise conclusions could be done using data from laboratory experiments on acoustic penetration or surface waves.

The principal conclusion of this section is that a detailed account of the physical nature of the sediment, compared to the simple viscous fluid model, provides some small correction in the penetration ratio. The main difference in the behavior of displacement profiles and of the reflection coefficient is in the vicinity of critical regime.

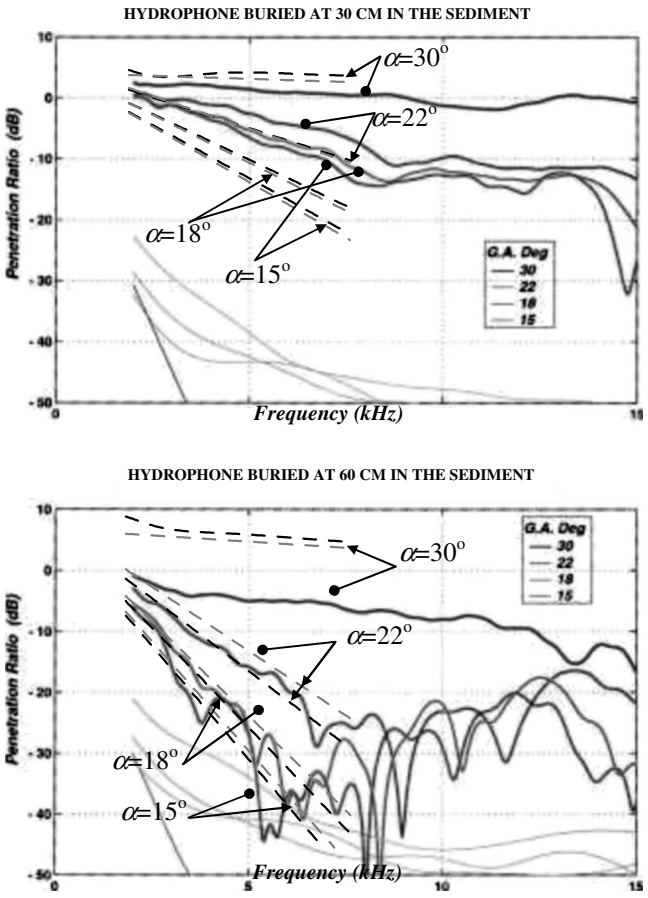


FIG. 3. Comparison of experimental penetration ratio (Ref. 14) (solid lines) and simulations for the Buckingham GS model (black dashed line for components $\sigma_{xx} \approx \sigma_{yy}$) and for the effective fluid (gray dashed lines) at frequencies 2–6 kHz. The upper plot is for 30 cm hydrophone depth, and the lower one is for 60 cm depth. The experimental data from Ref. 14 (solid lines) are reproduced with the permission of the Journal of the Acoustical Society of America.

III. GEOMETRICAL ASPECTS AND 2D PROBLEM

A. 2D geometry and surface roughness

The analysis presented in Sec. II shows that the oscillating (and even stochastic) behavior revealed experimentally[14,15] cannot be obtained using models that neglect surface or volume irregularities of sediment. Correspondingly, following Thorsos *et al.*[16] who used the exact representation[21] of the BEM,[22] we applied this formalism to the case of acoustic penetration into rough homogeneous fluid sediment. The original part of our analysis concerns a detailed characterization of cases when roughness enhances greatly the penetration mean level, as well as the statistical description of the problem that helps, for instance, to estimate the expected penetration level and its standard deviation in some given conditions.

The 2D geometry of the problem is presented in Fig. 4. A Gaussian beam with a focal point (x_b, y_b) and a focal width w illuminates a rough surface given as a random profile $y = \xi(x)$ having certain statistical properties. As has been suggested,[20] ripples in sand can be represented by the introduction of a random field $\xi(x)$ having the Gaussian spectrum $S(s)$ centered at the spatial frequency s_0 corresponding to the quasi-period $2\pi/s_0$ of ripples

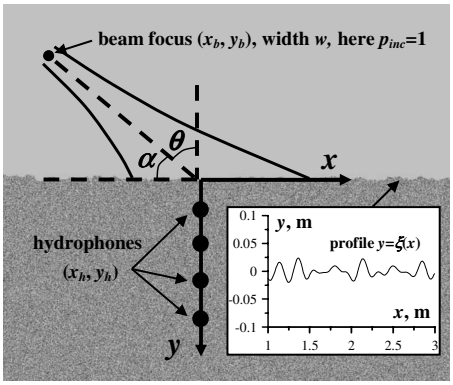


FIG. 4. Geometry of the 2D problem of beam penetration into rough sediment. The inset: a roughness realization generated with the given statistics.

$$S(s) = \begin{cases} e^{-\frac{(s-s_0)^2}{2s_w^2}} + e^{-\frac{(s+s_0)^2}{2s_w^2}}, & s > s_{\min} \\ 0, & s < s_{\min}, \end{cases} \tag{13}$$

with some low cut-off spatial frequency $s_{\min}$, which implies that the sediment is globally plane and horizontal, so that low-frequency waviness is absent. Here s_w is the spectrum width; for a smaller width the ripples are more equidistant. This spectrum should be multiplied by a complex random factor and Fourier-transformed in order to obtain a roughness realization $\xi(x)$ as follows:

$$\xi(x) = C\int_{-\infty}^{+\infty} \sqrt{S(s)}(r_1(s) + ir_2(s))\exp(isx)ds, \tag{14}$$

where $r_1(s)$ and $r_2(s)$ are independent random numbers having Gaussian distribution with zero average.

Use of a symmetrical spectrum $S(-s)=S(s)$ together with the requirements $r_1(s)=r_1(-s)$ and $r_2(s)=-r_2(-s)$ ensures real values of the profile $\xi(x)$. Each realization $\xi(x)$ can be normalized by fixing the constant C in Eq. (14) so that its root mean square equals a given rms value. Such a realization is plotted in the inset in Fig. 4 for rms=1.5 cm, $s_0 = 2\pi \cdot 3$ m^{-1}, $s_w = 2\pi \cdot 2$ m^{-1}, and $s_{\min} = 2\pi \cdot 2$ m^{-1}. This example will be used everywhere in our simulations.

B. Equations of the BEM

As it can be shown[21] using the second Green's identity, the acoustical pressures p_w and p satisfying the Helmholtz equation in water and in sediment, respectively, are expressed via the following integral representation:

$$p_w(x,y) = p_{\text{inc}}(x,y) + \int_{-\infty}^{+\infty} dx' \sqrt{1+\left(\frac{d\xi}{dx'}\right)^2} \left(\frac{\partial G_w}{\partial n'} p_w(x',y') - \frac{\partial p_w(x',y')}{\partial n'} G_w\right)\Bigg|_{y'=\xi(x')},$$

$$p(x,y) = -\int_{-\infty}^{+\infty} dx' \sqrt{1+\left(\frac{d\xi}{dx'}\right)^2} \left(\frac{\partial G}{\partial n'} p(x',y') - \frac{\partial p(x',y')}{\partial n'} G\right)\Bigg|_{y'=\xi(x')}, \tag{15}$$

where G is Green's function in 2D

$$G(x,y,x',y') = \tfrac{1}{4} i H_0^{(1)}(k\sqrt{(x-x')^2+(y-y')^2}), \tag{16}$$

in which $H_0^{(1)}$ is the Hankel function of the first kind, and k is the complex wavenumber in the sediment, $k=(\omega/c)(1+ia/54.58)$, with a the attenuation coefficient of the sediment, expressed in dB/λ. Green's function G_w in water can be obtained from Eq. (16) by substituting k with $k_w=\omega/c_w$. Here ω is the wave frequency, and c and c_w are the sound velocities in sediment and in water, respectively.

Equation (15) must be supplemented with the boundary conditions

$$p_w(x,y)|_{y=\xi(x)} = p(x,y)|_{y=\xi(x)},$$

$$\frac{1}{\rho_w}\frac{\partial p_w(x,y)}{\partial n}\Bigg|_{y=\xi(x)} = \frac{1}{\rho}\frac{\partial p(x,y)}{\partial n}\Bigg|_{y=\xi(x)}, \tag{17}$$

valid for a fluid-fluid interface of an arbitrary profile $\xi(x)$, with ρ_w and ρ, densities of water and of sediment, respectively. Here $\partial/\partial n$ denotes the derivative taken along the normal vector drawn at the point $(x, y=\xi(x))$ of the surface and directed upward as follows:

$$\frac{\partial}{\partial n} = \left(\frac{d\xi}{dx}\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right)\left(1+\left(\frac{d\xi}{dx}\right)^2\right)^{-1/2}. \tag{18}$$

The incident field $p_{\mathrm{inc}}(x,y)$ can be specified as the Gaussian beam

$$p_{\mathrm{inc}}(x,y) = \frac{1}{2\sqrt{\pi}} k_w w \int_{-\pi/2}^{\pi/2} e^{-(k_w w/2)^2(\theta'-\theta)^2} \times e^{ik_w((x-x_b)\sin\theta' + (y-y_b)\cos\theta')} d\theta' \tag{19}$$

represented as the plane-wave decomposition, with w having the sense of the beam width.[21] Note that the pressure in the focal point $p_{\mathrm{inc}}(x_b,y_b)=1$. The incident angle θ is shown in Fig. 4.

Equation (15) presents the solution for the pressure field in the entire space only if this field and its normal derivative are given at the surface. To find them, using the same Green's second identity at the boundary one obtains the system of integral equations

$$\frac{1}{2}p(x,y)\Bigg|_{y=\xi(x)} = p_{\mathrm{inc}}(x,y)|_{y=\xi(x)} + \int_{-\infty}^{+\infty} dx' (\partial' G_w p_w(x',y') - \partial' p_w G_w)|_{y'=\xi(x')},$$

$$\frac{1}{2}p(x,y)\Bigg|_{y=\xi(x)} = -\int_{-\infty}^{+\infty} dx' \left(\partial' G p_w(x',y') - \frac{\rho}{\rho_w}\partial' p_w G\right)\Bigg|_{y'=\xi(x')}, \tag{20}$$

where

$$\partial' = \left(\frac{d\xi}{dx'}\frac{\partial}{\partial x'} - \frac{\partial}{\partial y'}\right). \tag{21}$$

This system is discretized as proposed earlier[21] and then solved numerically.

In our simulations, for each set of parameters α, ω, w, and (x_b,y_b) defining the size of the insonified zone at the interface, the results generated by the BEM code in the absence of roughness were compared to an exact solution, which is easy to derive by replacing the incident plane wave in the integrand in Eq. (19) with analogous expressions for the reflected and the transmitted waves.[1] In practice, reducing the grazing angle α below 7°–10° for frequencies higher than about 8–10 kHz and below 10°–15° for frequencies lower than about 5–8 kHz is not recommended, since, for $\alpha\to 0$, the insonified zone formally becomes infinitely large. Because of diffractional divergence, these limits for α are even stricter if the focal point (x_b,y_b) moves away from the surface, especially at low frequencies.

In Sec. III C, we give calculation results for the acoustical pressure fields in various penetration regimes for single roughness realizations, describe frequency dependencies for pressure at various depths in sediment, and finally discuss some statistical properties of the calculated pressure fields.

C. Results of 2D modeling and discussion

The results discussed in this section have been obtained with the following values of physical constants: water density $\rho_w=1024$ kg/m^3, sediment density $\rho=1920$ kg/m^3, sound velocity in water $c_w=1515$ m/s and $c=1680$ m/s in sediment, zero attenuation in water, and sediment attenuation $a=0.35$ dB/λ. The parameters c and a approximately correspond to the values at 20 cm depth of the profiles $c_L(y)$ and $a_L(y)$ in the GS model.

Grayscale plots for the pressure field in the absence and in the presence of roughness are given in Figs. 5(a)–5(d) together with the pressure field sections at 0.1–0.4 m depths, for the beam parameters $x_b=y_b=0$ and $w=0.8$ m. Two angles $\alpha=15°$ and 30° lower and higher than critical and two frequencies $f=5$ and 15 kHz are considered. Black in the grayscale plots corresponds to maximum pressure amplitude for both plane and rough interface; this maximum is different in different sets (a)–(d). As expected, roughness produces higher effect at higher frequencies. This difference is less remarkable for grazing angles higher than critical and more important for subcritical angles, especially for high depths of about 0.4 m where a huge gain in penetration can be observed [see two lower sets in Fig. 5(c)]. A closer look at the pressure field images at high frequencies (8 kHz and more) demonstrates that pressure amplitude has strong maxima just below increasing fronts $\xi'(x)>0$ in the roughness profile:

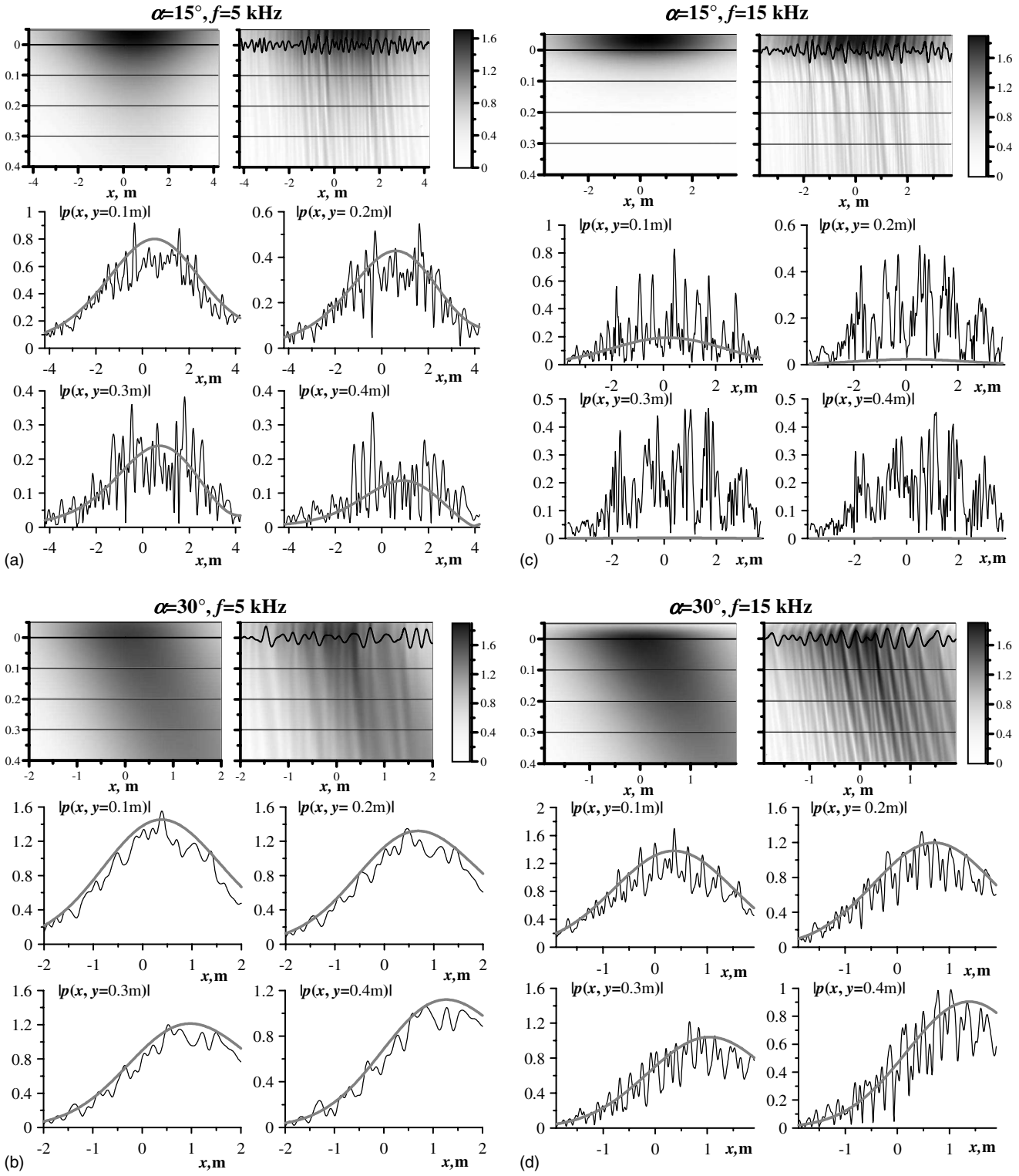


FIG. 5. Grayscale plots for the normalized [so that $p_{\mathrm{inc}}(x_b, y_b)=1$] pressure fields in the absence (exact solution for the flat interface, left figure) and in the presence (BEM simulation, right figure) of roughness and sections of these fields at depths $y=0.1-0.4$ m: the BEM results (thin black lines) compared to the exact solution (thick gray line). (a) $\alpha=15°$, $f=5$ kHz; (b) $\alpha=30°$, $f=5$ kHz; (c) $\alpha=15°$, $f=15$ kHz; and (d) $\alpha=30°$, $f=15$ kHz.

The local grazing angle for such segments can be higher than critical thus greatly enhancing acoustic penetration. Indeed, in the considered case, the standard deviation of the roughness profile slope is about 7.5°, and the characteristic incidence angle range for the plane waves in decomposition equation (19) is about 2°, so that some exceptionally steep slopes can make the local grazing angle higher than the nominal critical angle of 25.6°. This simple mechanism explains previous observations[14,15] of anomalously high energy penetration into rough sediments. This effect is especially pronounced when the wavelength is much smaller then the quasi-period of ripples. In our case [Fig. 5(c)] the wavelength of about 0.1 m is comparable to the quasi-period $2\pi/s_0=0.33$ m, which assumes the interpretation of the

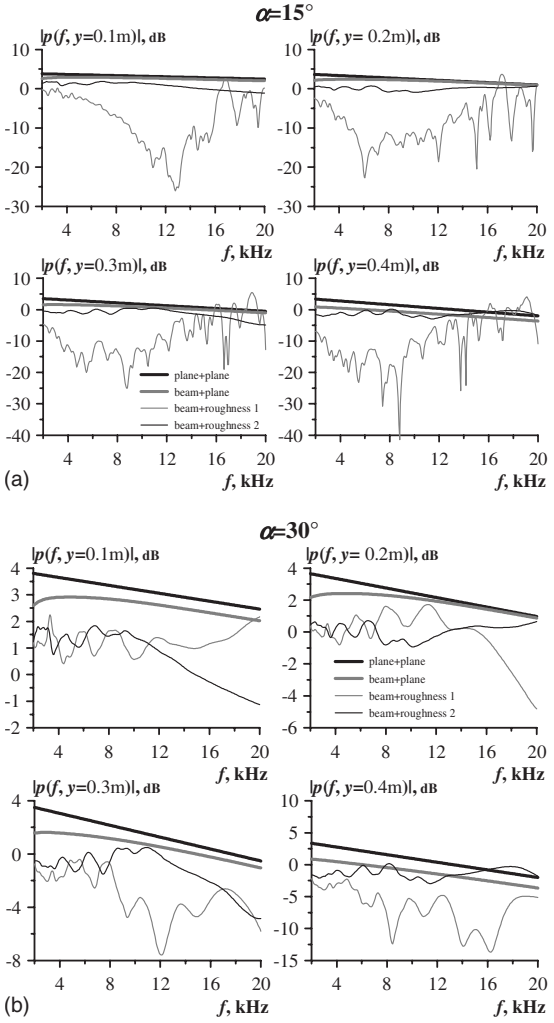


FIG. 6. The frequency dependence of the penetration ratio [the pressure amplitude in decibel related to the pressure at the beam focus $p_{\mathrm{inc}}(x_b, y_b) = 1$, see Eq. (19)] at different depths: the thick black line for the plane wave insonifying the plane interface, the thick gray line for the Gaussian beam and the plane interface, and the thin black and gray lines for the Gaussian beams and two realizations of roughness profile. (a) $\alpha = 15°$ and (b) $\alpha = 30°$.

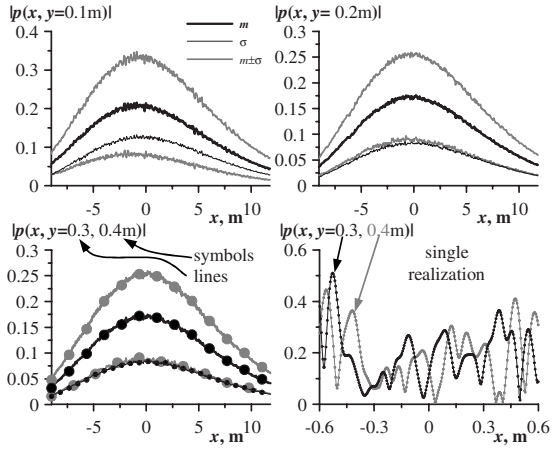


FIG. 7. Statistics of the pressure field for $f = 15$ kHz, $\alpha = 15°$ at depths $y = 0.1 - 0.4$ m. Two upper plots are for the depths 0.1 and 0.2 m: the statistical average and the standard deviation for 1000 roughness profile realizations (thick and thin black lines, respectively), and the average plus and minus the standard deviation (gray lines). The lower left plot: the lines represent the average, the standard deviation, and the average ± standard deviation at 0.3 m depth similarly to the upper set; symbols (large black for the average, small black for the standard deviation, and the gray ones for the average ± standard deviation) were added to represent statistics at 0.4 m depth. The lower right plot shows two pressure particular pressure fields for two single roughness realizations.

wave-interface interaction as a sort of diffraction. However, in any case, the anomalously high penetration level is explained in the model by the local tilting of the seafloor.

Frequency dependence of the penetration ratio[14] in dB is plotted in Figs. 6(a) and 6(b) for four different depths and two grazing angles $\alpha = 15°$ and 30°. The thick black line shows the results for plane waves and plane interfaces, the thick gray line illustrates the exact solution for plane interfaces, and two thin lines depict pressure amplitudes for two independent random realizations [Eq. (14)], for each of the angles. It is seen that the penetration problem has fully stochastic character, and that for subcritical angles there is an important gain for high frequencies (8–10 kHz and more). Even for low frequencies of about 2–4 kHz the roughness-induced change is substantial, so that one should not expect precise agreement between measurements[14] for a real seafloor and modeling for a plane interface (Fig. 3). Indeed, variations of the order of several dB appearing in Figs. 6(a) and 6(b) at frequencies 2–4 kHz are comparable with the mismatches between data and theoretical curves for a plane interface in Fig. 3. Other numerical experiments show that only at frequencies less than about 1 kHz the interfaces are viewed as "plane."

The stochastic character of the problem (see also paper[3] for the Monte Carlo simulations) deserves more close attention. A large number (1000) of realizations were generated for $f = 15$ kHz, $\alpha = 15°$ when the correlated component of the pressure field rapidly decays with depth because of low grazing angle, but the uncorrelated (roughness-induced) component remains present. Moderate diffraction divergence associated with the high frequency of 15 kHz enables us to consider a source located quite far away ($x_b = -50$ m) of the insonified zone, which is more realistic in the context of mine detection than the beam focused on the surface. Values of physical parameters are given in the beginning of this section, except the sediment attenuation $a = 0$ chosen in this case in order to separate the effect of material loss from the physics of acoustic penetration following subcritical penetration into the sediment. The upper left and right plots in Fig. 7 represent the average m pressure amplitude (thick black line) and the standard deviation σ (thin black line), together with their sum and difference $m \pm \sigma$ depicted in thick gray line, for depths of 0.1 and 0.2 m. The lower left plot contains the same values at 0.3 (lines) and 0.4 m depths (symbols): The respective curves for these two depths practically coincide in the considered case of lossless sediment. It is interesting to note that this depth independence appears below some minimum depth (for instance, the curves at 0.2 m slightly deviate) where the correlated component has already disappeared, and takes place for statistical characteristics only. Indeed, single realizations do not possess this property: Pressure curves shown in the lower right set in Fig. 7 are different for different depths. In addition, in the considered

regime without correlated pressure component, i.e., when only scattered field is present, the average m coincides with $m-\sigma$. Other numerical experiments have demonstrated that the properties $m \approx 2\sigma$ and $m \approx \text{const}(y)$ of the scattered field hold for various combinations of parameters. Theoretical interpretation of the observed empirical dependencies is the subject of our future research.

IV. CONCLUSIONS

In this paper, two approaches to acoustic penetration problem into rough sediments were considered: The first one uses an advanced constitutive model sacrificing geometrical irregularities, while the second one, in contrast, takes into account bottom roughness neglecting the complicated constitutive behavior. The first approach using the GS model describes well the geoacoustical properties of the sandy sediment considered here, as well as the slopes of the experimental penetration ratio curves;[14] their shifts are, however, poorly reproduced. This discrepancy is explained via the second approach that consists in application of the virtually exact BEM to penetration of an acoustic beam into a sediment with a rough interface. Our numerical analysis demonstrated that such a problem has stochastic nature even for low frequencies of about 5 kHz. For higher frequencies the impact of roughness is much more important and, at low grazing angles, provides a huge gain in penetration.

Statistical analysis of the scattered field has revealed some interesting empirical properties: The average frequently equals two standard deviations and does not depend on depth for lossless sediment. This enables us to hope that theoretical interpretation of such behavior can provide a more compact description of acoustical scattering by seafloors.

ACKNOWLEDGMENTS

This work was funded by the DGA (the Thales Underwater Systems subcontract). The authors are grateful to P. Penven, L. Leviandier, and N. Burlet for stimulating discussions.

[1]L. Brekhovskikh and Yu. Lysanov, *Fundamentals of Ocean Acoustics* (Springer-Verlag, New York, 1982).
[2]N. P. Chotiros, "Biot model of sound propagation in water-saturated sand," J. Acoust. Soc. Am. **97**, 199–214 (1995).
[3]E. I. Thorsos, D. R. Jackson, J. E. Moe, and K. L. Williams, "Modeling of subcritical penetration into sediments due to interface roughness," in *High Frequency Acoustics in Shallow Water*, edited by N. G. Pace, E. Pouliquen, O. Bergem, and A. P. Lyons (Nato Saclant, La Spezia, Italy, 1997).
[4]C. J. Hickey and J. M. Sabatier, "Choosing Biot parameters for modeling underwater sand," J. Acoust. Soc. Am. **102**, 1480–1484 (1997).
[5]M. J. Buckingham, "Compressional and shear wave properties of marine sediments: Comparison between theory and data," J. Acoust. Soc. Am. **117**, 137–152 (2005).
[6]M. Buckingham, "Wave propagation, stress relaxation, and grain-to-grain shearing in saturated, unconsolidated marine sediments," J. Acoust. Soc. Am. **108**, 2796–2815 (2000).
[7]J. D. Goddard, "Nonlinear elasticity and pressure-dependent wave speeds in granular media," Proc. R. Soc. London, Ser. A **430**, 105–131 (1990).
[8]V. Aleshin, V. Gusev, and V. Tournat, "Acoustic modes propagating along the free surface of granular media," J. Acoust. Soc. Am. **121**, 2600–2611 (2007).
[9]X. Jacob, V. Aleshin, V. Tournat, P. Leclaire, W. Lauriks, and V. E. Gusev, "Acoustic probing of the jamming transition in an unconsolidated granular medium," Phys. Rev. Lett. **100**, 158003 (2008).
[10]M. Wyart, L. Silbert, S. Nagel, and T. Witten, "Effects of compression on the vibrational modes of marginally jammed solids," Phys. Rev. E **72**, 051306 (2005).
[11]K. Walton, "The effective elastic moduli of a random packing of spheres," J. Mech. Phys. Solids **35**, 213–226 (1987).
[12]A. C. Kibblewhite, "Attenuation of sound in marine sediments: A review with emphasis on new low-frequency data," J. Acoust. Soc. Am. **86**, 716–738 (1989).
[13]D. J. Wingham, "The dispersion of sound in sediment," J. Acoust. Soc. Am. **78**, 1757–1760 (1985).
[14]A. Maguer, W. Fox, H. Schmidt, E. Pouliquen, and E. Bovio, "Mechanisms for subcritical penetration into a sandy bottom: Experimental and modeling results," J. Acoust. Soc. Am. **107**, 1215–1225 (2000).
[15]A. Maguer, E. Bovio, W. Fox, and H. Schmidt, "In situ estimation of sediment sound speed and critical angle," J. Acoust. Soc. Am. **108**, 987–996 (2000).
[16]E. Thorsos, D. Jackson, and K. Williams, "Modeling of subcritical penetration into sediments due to interface roughness," J. Acoust. Soc. Am. **107**, 263–277 (2000).
[17]D. R. Jackson, K. L. Williams, E. I. Thorsos, and S. G. Kargl, "High-frequency subcritical acoustic penetration into a sandy sediment," IEEE J. Ocean. Eng. **27**, 346–361 (2002).
[18]B. A. Kasatkin, "Anomalous phenomena in sound propagation near the sea floor: A review," Acoust. Phys. **48**, 379–387 (2002).
[19]H. Schmidt, OASES Version 3.1, User Guide and Reference Manual, Department of Ocean Engineering, MIT, 2004. http://acoustics.mit.edu/faculty/henrik/oases.pdf (Last viewed 1/28/2009).
[20]E. Pouliquen, A. P. Lyons, and N. G. Pace, "Penetration of acoustic waves into rippled sandy seafloors," J. Acoust. Soc. Am. **108**, 2071–2081 (2000).
[21]A. A. Maradudin, T. Michel, A. R. McGurn, and E. R. Mendez, "Enhanced backscattering of light from a random grating," Ann. Phys. **203**, 255–307 (1990).
[22]L. C. Wrobel and M. H. Aliabadi, *The Boundary Element Method* (Wiley, New Jersey, 2002).
[23]E. L. Hamilton, "Sound velocity gradients in marine sediments," J. Acoust. Soc. Am. **65**, 909–922 (1979).
[24]E. L. Hamilton, "Shear-wave velocity versus depth in marine sediments: A review," Geophysics **41**, 985–996 (1976).
[25]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).
[26]S. D. Conte and C. deBoor, *Elementary Numerical Analysis* (McGraw-Hill, New York, 1972).

Under-ice noise generated from diamond exploration in a Canadian sub-arctic lake and potential impacts on fishes

D. Mann[a)]
College of Marine Science, University of South Florida, 140 Seventh Avenue South, St. Petersburg, Florida 33701-5016

P. Cott
Department of Fisheries and Oceans, 101-5204-50th Avenue, Yellowknife, Northwest Territories X1A 1X1, Canada

B. Horne
AMEC Earth and Environmental, 2227 Douglas Road, Burnaby, British Columbia V5C 5A9, Canada

(Received 1 January 2009; revised 13 July 2009; accepted 15 July 2009)

Mineral exploration is increasing in Canada, particularly in the north where extensive diamond mining and exploration are occurring. This study measured the under-ice noise produced by a variety of anthropogenic sources (drilling rigs, helicopters, aircraft landing and takeoff, ice-road traffic, augers, snowmobiles, and chisels) at a winter-based diamond exploration project on Kennady Lake in the Northwest Territories, Canada to infer the potential impact of noise on fishes in the lake. The root-mean-square noise level measured 5 m from a small diameter drill was approximately 46 dB greater (22 kHz bandwidth) than ambient noise, while the acoustic particle velocity was approximately 40 dB higher than ambient levels. The loudest sounds at the exploration site were produced by ice cracking, both natural and during landing and takeoff of a C130 Hercules aircraft. However, even walking on the snow above the ice raised ambient sound levels by approximately 30 dB. Most of the anthropogenic sounds are likely detectable by fishes with hearing specializations, such as chubs and suckers. Other species without specialized hearing adaptations will detect these sounds only close to the source. The greatest potential impact of noise from diamond exploration is likely to be the masking of sounds for fishes with sensitive hearing. [DOI: 10.1121/1.3203865]

PACS number(s): 43.30.Nb, 43.80.Nd, 43.50.Rq, 43.80.Lb [MCH] Pages: 2215–2222

I. INTRODUCTION

Mineral exploration and mining development are increasing in Canada, particularly in the north where extensive diamond mining and exploration are occurring (Birtwell *et al.*, 2005). These activities pose a multitude of potential adverse environmental impacts that must be addressed during the environmental impact review process (Erlandson and Associates, 2000; Cott *et al.*, 2003). One potential effect, commonly brought forward during public consultations, is the potential impact of anthropogenic noise, such as from exploratory drilling and ice-road traffic, on fishes (Stewart, 2001). Therefore, in order for proponents and regulators to evaluate the effect of noise on fishes during project evaluation, four pieces of information are necessary: (1) the frequency and intensity of noise created underwater by different project activities, (2) the audiograms of fishes in the vicinity of the proposed project, (3) the propagation of the sounds, and (4) an assessment of the potential impact of the noise on the fishes that may be present (physical and/or behavioral impairments) that could affect the survival of individual fishes exposed. In Canada, the killing of fishes by means other than fishing is prohibited under the federal Fisheries Act (Government of Canada, 1985). A proponent, therefore, has the potential to contravene the Fisheries Act if underwater noise from their proposed project results in fish mortality.

The effects of noise on the hearing and behavior of fishes are poorly understood; however, there is evidence that anthropogenic noise can have effects on fishes (McCauley *et al.*, 2003; Popper *et al.*, 2005). It may be possible to predict the effect that noise has on a fish by determining the levels and frequencies of the noise and comparing those to the audiogram of fishes in the receiving environment (Popper *et al.*, 2005).

The purpose of this paper is to characterize underwater sounds produced by a variety of anthropogenic sources common to a winter-based diamond exploration project on a frozen lake in the Northwest Territories, Canada. These sources included ice-road traffic, aircraft landings, movement of heavy equipment, and exploratory drilling. These noise sources are also found in other types of northern development such as hydrocarbon exploration projects and base-metal mining. When coupled with information on the hearing capabilities of northern fishes, this information can be used to help assess potential effects of noise production on the hearing of fishes and to assist in the development of mitigation measures to minimize or avoid impacts.

II. METHODS

This study was conducted at Kennady Lake, located approximately 300 km northeast of Yellowknife, Northwest

[a)]Author to whom correspondence should be addressed. Electronic mail: dmann@seas.marine.usf.edu

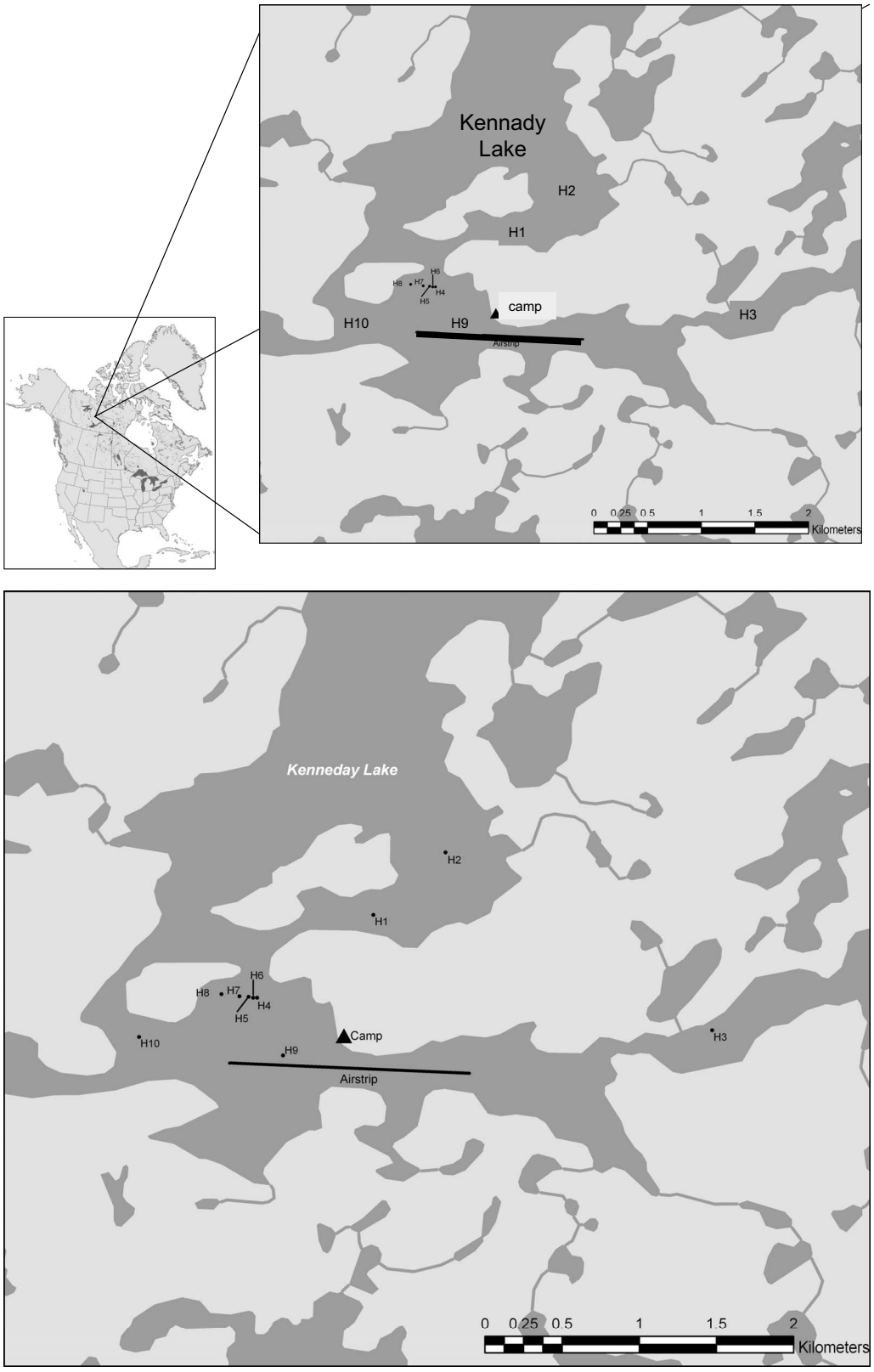


FIG. 1. Location of the Gacho Kué diamond exploration site, 300 km northeast of Yellowknife, Northwest Territories, Canada. Sites of recordings are indicated by H (for hole drilled through the ice).

Territories, Canada and approximately 20 km north of the tree-line in the sub-arctic tundra (Fig. 1). Kennady Lake is the site of the Gahcho Kué diamond exploration project operated by DeBeers Canada. Access to the site is by aircraft or by a 121 km ice-road branching off of the 568 km Tibbitt-Contwoyto winter road that leads north to the Ekati and Diavik diamond mines, 495 km of which crosses frozen lakes.

Kennady Lake is an 814 ha headwater lake in the Lockhart River drainage and drains into Great Slave Lake (Fig. 1). Kennady Lake is an oligotrophic lake typical of many

TABLE I. Summary of acoustic measurements made at the Gahcho Kué diamond exploration site, Northwest Territories, Canada. Measurements were made at a depth of 6 m. Sound pressure levels marked with * indicate a bandwidth from 2 to 10 000 Hz; otherwise, bandwidths are 0–22 050 Hz. The site of each recording is indicated in the first column. Data are sorted by increasing band sound pressure level noise in the 200–300 Hz band.

Noise source	Hole no.	Water depth (m)	Ice thickness (cm)	Source distance (m)	Peak frequency (Hz)	rms SPL (dB re 1 μPa)	Peak SPL (dB re 1 μPa)	Band level (200–300 Hz) (dB re 1 μPa)
Ambient	H3	7.5	175	0	45	82.3	114.4	64.2
Snowmobile idling	H3	7.5	175	0.5	29	110.2	119.5	78.5
Walking	H3	7.5	175	0.1	4,326	111.3	130.5	86.8
Helicopter idling	H3	7.5	175	1	83	116.4	130.6	89.9
Pickup truck	H9	7.0	166	30	1627	107.0*	149.7	92.3
Helicopter landing	H3	7.0	175	1	39	118.0	148.8	93.3
Gas auger	H3	7.0	175	0.5	76	125.5	149.2	96.0
Helicopter hovering	H3	7.0	175	50	83	119.4	150.1	96.4
Snowmobile, 40 km/h	H3	7.0	175	0.5	87	120.1	148.7	98.0
Snowmobile, 20 km/h	H3	7.0	175	0.5	99	121.4	146.7	98.3
Snowmobile, 60 km/h	H3	7.0	175	0.5	75	119.9	147.7	102.6
Dump truck (empty)	H10	6.7	166	30	121	114.8*	153.6	107.1
Dump truck (loaded)	H10	6.7	166	30	237	115.5*	154.5	109.2
Ice chisel	H3	7.0	175	0.5	113	133.2	148.6	112.8
Large casing drill	H1	9.0	236	15	183	124.5	148.4	113.5
Small diameter coring drill	H4	14.6	163	5	88	127.8	141.8	115.6
C130-H takeoff	H9	7.0	166	260	67	>126.9	>146.9	>116.3
Ice cracking	H9	7.0	166	105 (estimated)	120	131.4	145.8	118.8
Ice cracking	H9	7.0	166	Unknown	171	>134.8	>146.4	>124.6
C-130 Hercules aircraft landing	H9	7.0	166	260	318	>136.1	>147.4	>125.3
Grader	H10	6.7	166	10	179	>136.4	>149.6	>129.4

sub-arctic lakes in the Canadian Precambrian Shield and is comprised of three main basins, with a mean depth of approximately 5 m and a maximum depth of 18 m. The lake is covered by ice for 7–8 months of the year, with maximum ice thickness of approximately 2 m. The fish community is simple and includes lake trout (*Salvelinus namaycush*), lake chub (*Couesius plumbeus*), Arctic grayling (*Thymallus arcticus*), northern pike (*Esox lucius*), burbot (*Lota lota*), nine-spine stickleback (*Pungitius pungitius*), slimy sculpin (*Cottus cognatus*), and round whitefish (*Prosopium cylindraceum*). Round whitefish are the most abundant species, while lake trout are the top predators (DeBeers Canada, 2005).

The study was conducted between March 24 and 27, 2006 and included under-ice measurements of noise from a variety of sound sources common to mineral exploration at various distances. These sound sources included a *small diameter coring drill* (6.35 cm), *large diameter casing drill* (installing a 0.66 m casing inside a pre-drilled 0.763 m casing with three compressors running), *snowmobile* (Bombardier Skandic Ski-Doo with a Rotax 440F engine), *ice auger* (Husqvarna 227 gas-powered auger with a 20 cm cutting diameter), *ice chisel*, *humans walking*, *large fixed-wing aircraft* (Hercules C-130H; ~73 625 kg), *pick-up truck*, *gravel truck*, *grader* (first gear moving at 10 km/h with the blade, wing, and plow down), and *helicopter* (Bell 206 Jet Ranger). The locations of each monitoring site with reference to the lake and camp infrastructure are shown in Fig. 1 and Table I. Water temperature ranged from 0.75 °C near the surface to 3.61 °C at the bottom.

An HTI-96-min hydrophone (sensitivity −164 dB re 1 V/μPa; 2 Hz–30 kHz) (High-Tech Inc., Gulfport, MS) was used to measure acoustic pressure. An acoustic particle velocity probe and a geophone (sensitivity of 9.36 mV/cm/s and bandwidths of 10 Hz–1 kHz) (Acoustech Corporation, Philadelphia, PA) were suspended inside an aluminum frame which was mounted on a telescopic rod to measure acoustic particle velocity for the small diameter drill. At each sample site, a 20 cm hole was augered through the ice, and the hydrophone and/or geophone were lowered to make the recordings. Distances from the sound sources were measured with a measuring tape or laser range finder depending on distance. Noise was recorded under the ice surface at various depths and at various distances from the sound sources with a hydrophone. Particle velocity measurements were taken near the small diameter drill and at a control site (H3) (Fig. 1). Recordings were made with the sensor 10 cm above the bottom (with the frame resting on the bottom to avoid movement from the rod) with the geophone mounted vertically and then with the geophone mounted in the horizontal orientation directed toward the small diameter drill. Recordings were not made with the particle velocity meter at the other sites because the suspension of the probe failed from the cold water after the H7 horizontal measurement.

Ambient sound levels were measured in the easternmost basin of Kennady Lake at site H3, which is the basin furthest from exploration activities (Fig. 1). Diamond exploration activities that occurred on the lake during measurement of ambient sound levels included snowmobile, truck, heavy equipment traffic, and two active drilling small-core drill rigs. However, none of these activities occurred closer than 4 km from the ambient sound sample location when measurements were taken.

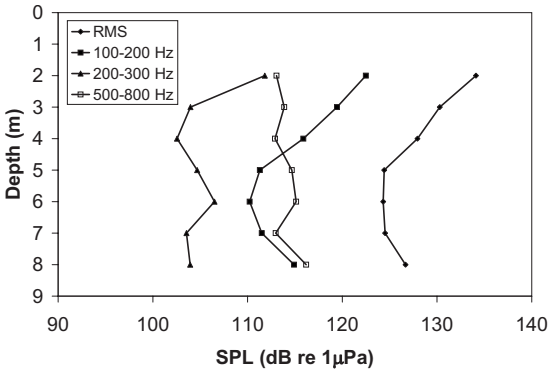


FIG. 2. Noise levels from casing rig at hole H1. rms sound pressure level (dB re 1 μPa) measured over the entire frequency band (2–2050 Hz) and band sound pressure level measured over three frequency bands.

The signals were recorded on handheld Pocket PC computers (Toshiba E755 and Dell Axim X50) with LOGGERHEAD software (Loggerhead Instruments, Inc., Sarasota, FL). The sample rate was either 44 100 Hz (Toshiba) or 22 050 Hz (Dell), and each recorder had a built-in anti-aliasing filter. Spectral analysis was performed using the fast Fourier transform (FFT) functions in MATLAB (Mathworks Inc., MA) with custom software (QLOGGER and QWAVFFT, D. Mann). Particle velocity (vertical and horizontal) and pressure measurements were conducted sequentially (separated by <5 min), rather than simultaneously, because the recorder was a single channel device. The duration of the analysis for sounds were 0.1 s for a single ice chisel strike, 14 s for the snowmobile passes, 5 s for walking, 7.3 s for the ice auger, 10 s for the trucks passing the hydrophone, 10 s for the Hercules C-130 aircraft landing and takeoff, 100 s for helicopter hovering, 30 s for helicopter landing, and 78 s for helicopter running on the ice. The duration was chosen to represent periods when most of the acoustic energy was present. Thus, it is much shorter for short-duration events such as ice chisel strikes than events that have a longer time course such as a helicopter hovering. For each sound analysis segment, the root-mean-square (rms) and maximum peak sound pressure levels were calculated from the time domain signal. The peak frequency and band pressure level sound over the 200–300 Hz band were calculated because this band is where most local fishes investigated have their most sensitive hearing (Mann *et al.*, 2007). The entire signals as described above were used in the FFT calculations (MATLAB), so the number of points in the FFT was the duration of the signal segment (s) multiplied by the sample rate (Hz).

III. RESULTS

Recordings made 15 m from the drilling of the large diameter casing at H1 showed consistent levels regardless of depth with most energy at low frequencies (Fig. 2). Recordings of the small diameter drill showed a similar pattern with little dependence on depth of the recording [Fig. 3(a)]. Sound levels decreased with distance in a manner that was intermediate between spherical and cylindrical spreading loss models at distances up to 120 m but decreased by a larger amount to the 240 m site [Fig. 3(b)]. Both acoustic pressure and

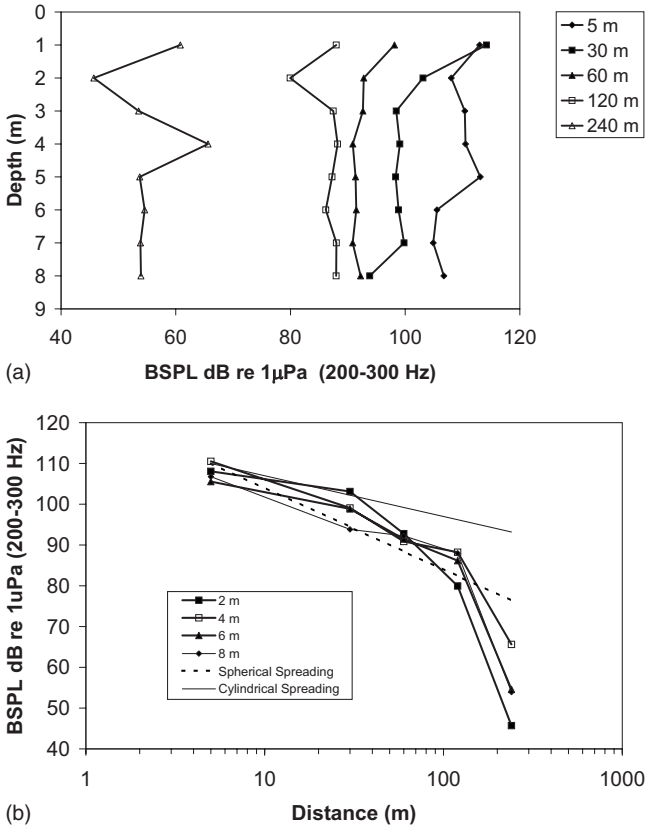
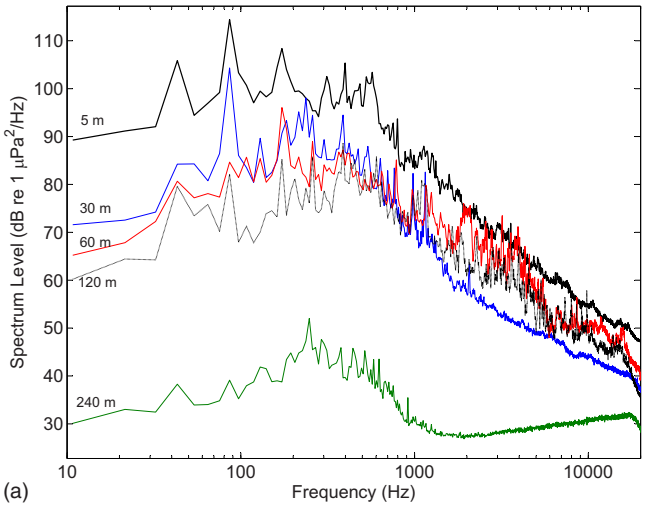


FIG. 3. Small diameter drill band sound pressure level in the 200–300 Hz frequency band. (a) Data plotted as a function of hydrophone depth at different distances from the small diameter drill. Legend shows distance of hydrophone from drill. (b) Data plotted as a function of the hydrophone distance from the drill. Theoretical attenuation from spherical and cylindrical spreading loss models is shown for comparison.

acoustic particle velocity had the greatest energy at the lowest frequencies (Fig. 4 and Table II). The acoustic particle velocity measurements were largely similar in the vertical and horizontal directions (Table II). However, because the recordings could not be made simultaneously in both directions with the probe, direct comparisons are not possible. The acoustic particle velocity measured 5 m from the small diameter drill was approximately 40 dB higher than the control site (H3) (Table II).

Ambient noise was recorded for an extended period at H2 (Fig. 5). These recordings showed that sound levels tended to stay at around 83 dB re 1 μPa, with a few periods of higher level sounds, which were associated with snowmobiles and compressor operations near the large diameter drill (hole H1).

The drilling operations are supported by a variety of equipment that are common noise sources at the site. Recordings made of a helicopter showed similar levels when the helicopter was running at full power while sitting on the ice and when it was hovering 5 m above the ice with increases of about 60 dB relative to the ambient noise levels mostly below 200 Hz (Fig. 6). The snowmobile was loudest below 200 Hz when idling or traveling slowly over the site (Fig. 7). The snowmobile, ice auger, chisel, and walking noise also had a higher frequency component between 1000 and 10 000 Hz, which was associated with compression of 5



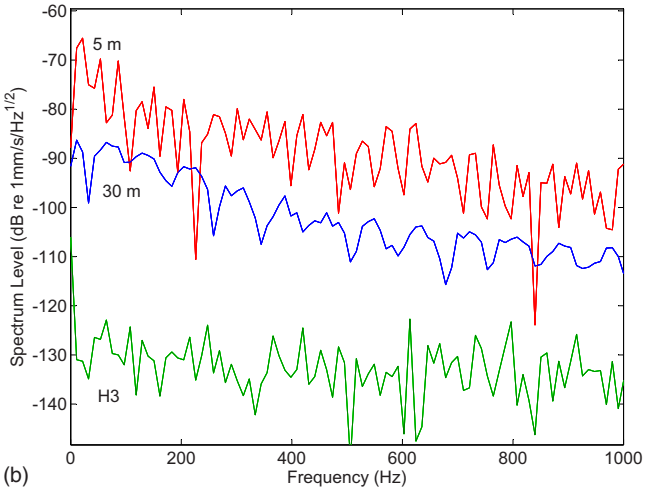


FIG. 4. (Color online) Small diameter drill acoustic spectra. (a) Spectrum level acoustic pressure at five distances at 6 m depth. (b) Spectrum level acoustic particle velocity measured in the vertical direction at 5 and 30 m from the small diameter drill and at a reference hole (H3).

cm of snow on top of the ice (Figs. 7 and 8). The C-130 Hercules aircraft landing produced the loudest sounds, which exceeded the recorder's maximum input level of 150 dB_{peak} re 1 μPa (Fig. 9). Most of these sounds were due to ice cracking during landing, which also happened naturally at

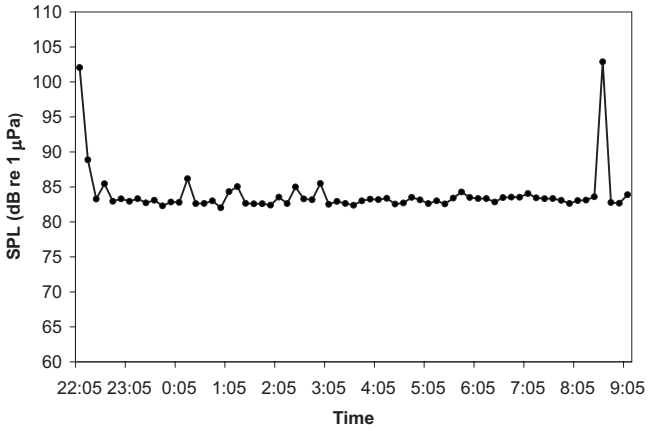


FIG. 5. Ambient noise (rms SPL, bandwidth 22 050 Hz) monitored at site H2 (dB re 1 μPa) over a 10.75 h period. Peaks near 22:00 and 08:30 correspond to snowmobile movements near the site.

other times. An ice crack that occurred naturally prior to the C-130 Hercules aircraft landing and which did not exceed the recorder's maximum input level is shown in Fig. 10 and Table I. In ice, the longitudinal propagating wave travels at approximately 3100 m/s, about twice as fast as the speed of sound in water (Stein, 1988). The arrival of the longitudinal wave from a single natural ice crack can be seen as the lower amplitude signal at 0.01 s in Fig. 10. The acoustic signal has a greater amplitude but is delayed relative to the longitudinal wave at 0.05 s. Assuming that the speed of sound in freshwater is 1420 m/s, the source of this ice crack was approximately 105 m away from the hydrophone.

The trucks produced sounds that were loudest from 100 to 1000 Hz (Fig. 11). The loudest truck was the grader, with lower sound levels from the dump truck and pickup truck (Fig. 11). The sound level was not very different for a full dump truck and an empty dump truck (Fig. 11).

The loudest sounds were produced by the grader and ice cracking, both natural and during landing and takeoff of a C-130 Hercules aircraft (Table I). The ice-cracking sounds produced signals that exceeded the recording range of the equipment, and thus their levels were reported as greater than the maximum level at which clipping occurred. Even simple

TABLE II. Acoustic particle velocity measurements made at different distances from the small diameter coring drill and at a control site (H3). Measurements were made in the vertical and horizontal planes. No vertical measurement is available for H7. The last column presents acoustic pressure measurements at the same location. Note that the particle velocity and pressure measurements were not made simultaneously but sequentially.

Site	Water depth (m)	Ice thickness (cm)	Distance from source (m)	Direction	rms velocity (dB re 1 mm/s)	Peak velocity (dB re 1 mm/s)	Band level (200–300 Hz) (dB re 1 mm/s)	rms SPL (dB re 1 μPa)
H3	7.5	175	Control	Vertical	−95.1	−83.3	−111.2	82.3
H3	7.5	175	Control	Horizontal	−93.3	−71.6	−113.3	82.3
H4	14.6	163	5	Vertical	−54.7	−35.0	−66.2	127.7
H4	14.6	163	5	Horizontal	−58.8	−44.3	−73.4	127.7
H5	11.0	150	30	Vertical	−79.3	−56.0	−76.1	113.1
H5	11.0	150	30	Horizontal	−63.9	−48.1	−74.8	113.1
H7	8.9	117	60	Horizontal	−74.9	−56.5	−87.8	113.2
H8	9.2	120	120	Vertical	−77.2	−52.3	−79.9	108.6
H8	9.2	120	120	Horizontal	−83.1	−66.5	−97.1	108.6

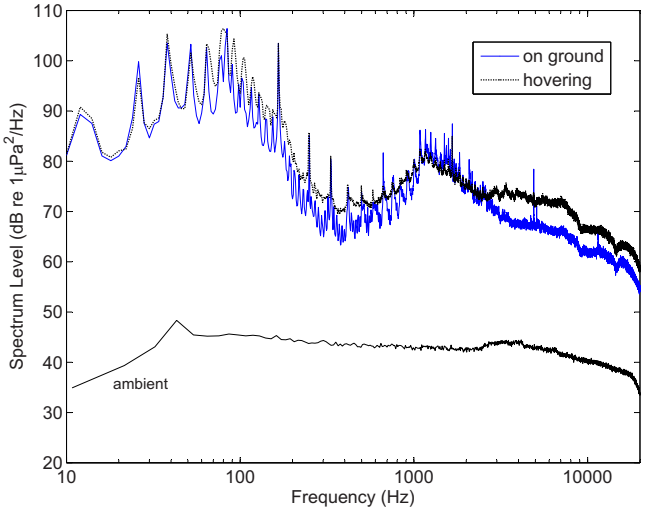


FIG. 6. (Color online) Spectrum level sound of a Bell 206 Jet Ranger helicopter hovering (dashed line) at 15 m and while touched down on ice (H3). Hovering simulated use of a 15 m long sling line used to transport supplies and equipment. Ambient noise shown for reference.

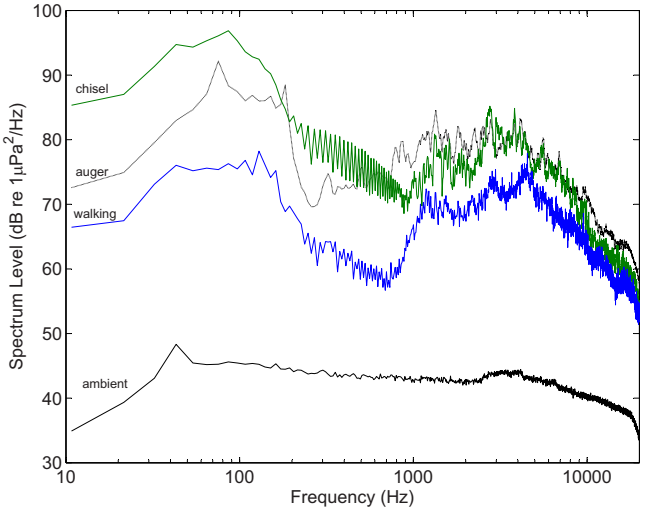


FIG. 8. (Color online) Spectrum level noise from a gas-powered ice auger (dashed line), hand ice chisel, and a person walking in comparison to ambient noise (H3).

activities such as walking on the snow above the ice raised ambient sound levels by approximately 30 dB (Table I).

IV. DISCUSSION

The natural background noise of this sub-arctic lake varies from very quiet (64.2 dB band level in the 200–300 Hz band) to relatively loud because of short-duration ice cracks (>125 dB band level in the 200–300 Hz band). All of the anthropogenic noise measured fell within this range of natural ambient noise. The main difference was the duration of the noise, which could be long, for example, in the case of compressors and drills that can run uninterrupted for hours. In contrast, the sounds from ice cracks last less than 1 s. The noise generated from ice cracking is impulsive in nature and can propagate both through the ice and through the underlying water and is an important contributor to background ambient noise both in lakes and in the Arctic Ocean (Milne, 1966; Stein, 1988).

The levels recorded from the small diameter drill (128 dB re 1 μPa at 5 m; dropping to approximately 80 dB re 1 μPa at 240 m) were quieter than those from a drilling facility located on the Beaufort Sea where drilling using much larger equipment produced background noise levels of 124 dB re 1 μPa at 1 km in a 10–10 000 Hz band (Blackwell *et al.*, 2004). Also, the particle velocity measurements just above the lake bed of Kennady Lake were much smaller in magnitude than ice-borne vibrations measured in the Blackwell *et al.* (2004) study. Despite the fact that ice and water have very different acoustic impedances, this is relevant for comparing potential particle motion sound exposures of a fish just below the ice to a fish on the bottom.

These acoustic measurements can be compared to fish audiograms to estimate the potential impacts of these noise sources on the fish fauna. It is important to note that for the

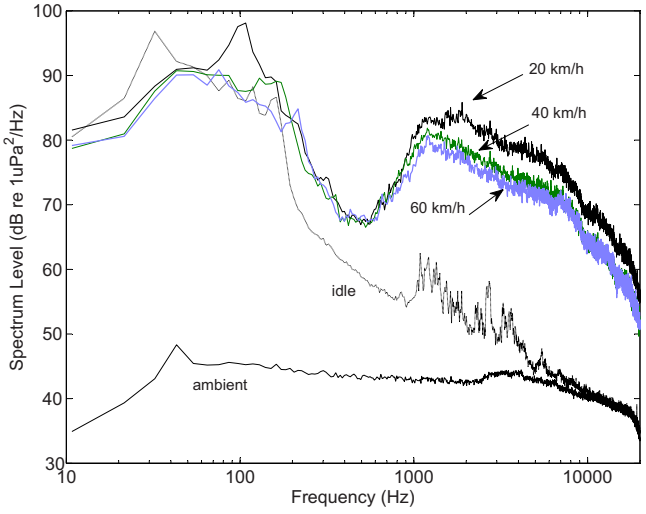


FIG. 7. (Color online) Spectrum level noise of snowmobile under four conditions: passing at 20, 40, and 60 km/h and idling (dashed line) above hydrophone (H3). Ambient noise is shown for reference.

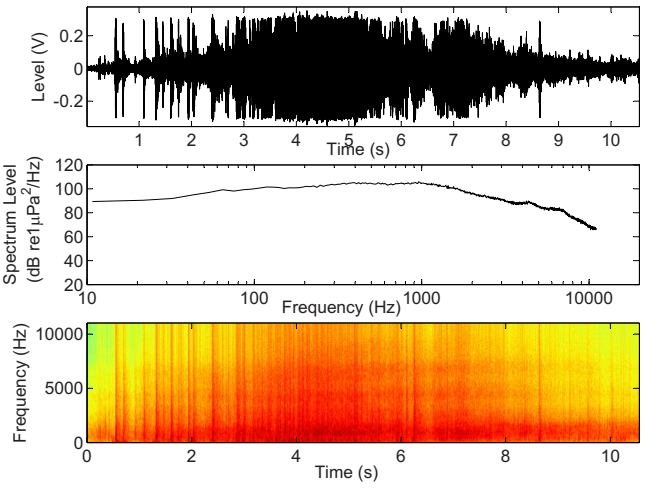


FIG. 9. (Color online) C-130 Hercules aircraft landing and resultant ice cracking monitored at hole H9. Note that ice cracking is clipped on the recorder.

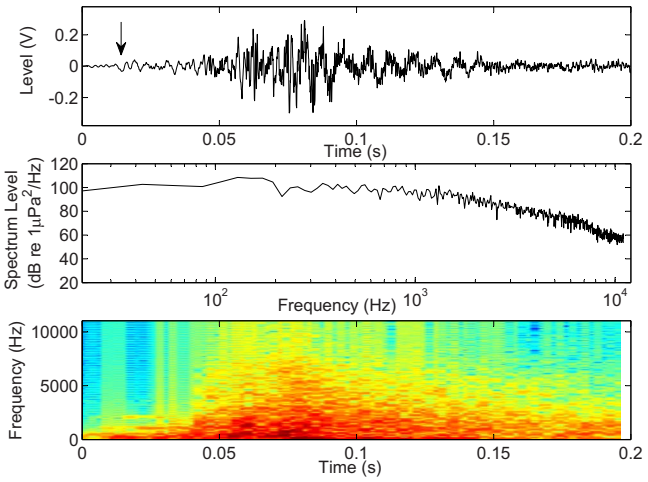


FIG. 10. (Color online) Single natural ice crack prior to C-130 Hercules aircraft landing at hole H9. Arrow indicates initial arrival through ice path in top plot.

fishes in Kennady Lake, hearing thresholds for some species have been measured using auditory evoked potential (AEP) techniques (Kenyon *et al.*, 1998). AEPs have been shown to yield similar results to behavioral methods in some species, but they have not been compared to behavioral measurements for these species. Thus, it is possible that the behavioral hearing thresholds of the fishes in Kennady Lake could be lower than the AEP thresholds. Kennady Lake has lake chub, which are known to have sensitive hearing (DeBeers Canada, 2005; Mann *et al.*, 2007). Lake chub (AEP) hearing thresholds at 200 Hz were 64 dB re 1 μPa (Mann *et al.*, 2007), which are similar to other otophysans, including goldfish, which have had hearing measurements made using behavioral techniques (Fay, 1988). Assuming a critical band of approximately 10% measured with behavioral techniques (Tavolga, 1974), which effectively lowers the threshold for detecting noise by 10 log (frequency), lake chub will likely be able to detect all of the anthropogenic sounds measured,

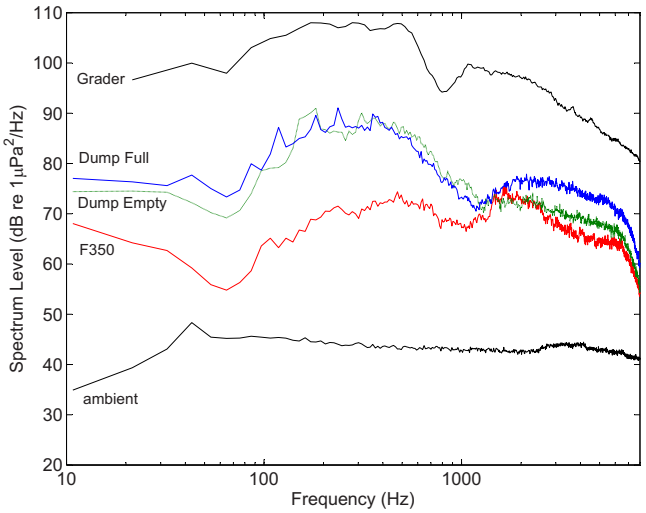


FIG. 11. (Color online) Spectrum level noise from road grader and three trucks driving on an ice-road past H10. Note that the grader was recorded at a sample rate of 44 100 Hz; the other trucks were recorded at 22 050 Hz. The ambient noise level is shown for reference.

including walking. Since band level drill sounds were about 116 dB re 1 μPa in the 200–300 Hz band at 5 m, they are about 45 dB above this hearing threshold of lake chub, which is likely loud enough to produce masking but not temporary hearing loss (also known as temporary threshold shift) (Fay, 1974; Smith *et al.*, 2004; Popper *et al.*, 2005; Wysocki and Ladich, 2005). AEP hearing thresholds of broad whitefish (*Coregonus nasus*), a species closely related to the round whitefish found in Kennady Lake, measured with auditory evoked potentials were 106 dB re 1 μPa at 200 Hz (Mann *et al.*, 2007), which suggests that they would be able to detect the drill sounds at 5 m.

The auditory evoked potential particle velocity threshold of broad whitefish was estimated to be −56 dB re 1 mm/s at 200 Hz; however, it is not possible to know whether the fish is sensitive to acoustic pressure, particle motion, or both given that the AEP measurements were made in a small tank (Mann *et al.*, 2007). Accurate particle velocity thresholds have been measured using a shaker system to simulate acoustic particle velocity with other fish species. While these fishes are not found in Kennady Lake, they allow the particle velocity measurements of the small diameter drill to be placed in perspective. The small diameter drill band level acoustic particle velocities measured at 5 m between 200 and 300 Hz were −66 dB re 1 mm/s in the vertical direction and −73 dB re 1 mm/s in the horizontal direction. Particle velocity measurements using a shaker system of the oscar (*Astronotus ocellatus*) had behavioral thresholds at 100 Hz from −62 to −60 dB re 1 mm/s (Lu *et al.*, 1996). Particle velocity measurements of sleeper goby (*Dormitator latifrons*) using a shaker had saccular afferent thresholds at 100 Hz ranging from −24 to −78 dB re 1 mm/s (Lu *et al.*, 1998). AEP thresholds of the sleeper goby were −58 dB re 1 mm/s at 100 Hz of, −56 dB re 1 mm/s at 200 Hz, and −59 dB re 1 mm/s at 345 Hz (Lu and Xu, 2002). These measurements are of the same magnitude as those obtained with AEPs with the broad whitefish (Mann *et al.*, 2007). Assuming that critical bandwidths for acoustic particle velocity are similar to critical bandwidths for acoustic pressure (i.e., band thresholds are 23 dB lower than tone thresholds at 200 Hz), this suggests that the particle velocity of the drill would be detectable but only at about 13 dB above threshold.

Anthropogenic sounds associated with an exploratory drilling operation can raise the under-ice background noise levels in a lake significantly. However, the main reason for this is that the natural ambient level is typically very low (except for ice cracks), so that even quiet sounds, such as those associated with walking on snow above the ice are easily detected. Most of these anthropogenic sounds are likely only detectable by fishes with hearing specializations, such as the chubs and suckers that possess Weberian ossicles. For species without specialized hearing adaptations, these sounds would likely be undetectable. The greatest potential impact of anthropogenic noise of the type measured in this study is likely to be the masking of natural sounds for fishes with the most sensitive hearing. While masking could be demonstrated, its effect on reproduction and survival is not known.

ACKNOWLEDGMENTS

The authors would like to thank M. Podolsky of DeBeers Canada and his staff at Gahcho Kué camp for the hospitality and assistance they received while on site. Assistance with study design and project logistics were provided by D. Balint, K. Cott, A. Demeule (Department of Fisheries and Oceans) and R. Johnstone (DeBeers Canada). P. Brunette (DFO) is acknowledged for map production Thanks are also extended to M. Somers and D. Miller (DFO), M. Hastings, and three anonymous reviewers whose comments greatly improved this paper.

Birtwell, I. K., Samis, S. C., and Kahn, N. Y. (**2005**). "Commentary on the management of fish habitat in northern Canada: Information requirements and policy considerations regarding diamond, oil sands and placer mining," Can. Tech. Rep. Fish. Aquat. Sci., Report No. 2607.

Blackwell, S. B., Greene, C. R., Jr., and Richardson, W. J. (**2004**). "Drilling and operational sounds from an oil production island in the ice-covered Beaufort Sea," J. Acoust. Soc. Am. **116**, 3199–3211.

Cott, P. A., Hanna, B. W., and Dahl, J. A. (**2003**). "Discussion on seismic exploration in the Northwest Territories 2000–2003," Can. Man. Rep. of Fish. and Aquat. Sci., Report No. 2648.

DeBeers Canada. (**2005**). "Application report for the Mackenzie Valley Land and Water Board," DeBeers Canada-Gahcho Kué Project, Mackenzie Valley Land and Water Board, Yellowknife, NWT, November.

Erlandson and Associates. (**2000**). "Oil and gas approvals in the Northwest Territories–Southern Mackenzie Valley," The Regulatory Roadmaps Project, Erlandson and Associates, Victoria, BC.

Fay, R. R. (**1974**). "Masking of tones by noise for the goldfish (*Carassius auratus*)," J. Comp. Physiol. Psychol. **87**, 708–716.

Fay, R. R. (**1988**). *Hearing in Vertebrates: A Psychophysics Databook* (Hill-Fay Associates, Winnetka, IL).

Government of Canada. (**1985**). Fisheries Act, c.F-14, amended list, April 1993.

Kenyon, T. N., Ladich, F., and Yan, H. Y. (**1998**). "A comparative study of hearing ability in fishes: The auditory brainstem response approach," J. Comp. Physiol., A **182**, 307–318.

Lu, Z., Popper, A. N., and Fay, R. R. (**1996**). "Behavioral detection of acoustic particle motion by a teleost fish (*Astronotus ocellatus*): Sensitivity and directionality," J. Comp. Physiol., A **179**, 227–233.

Lu, Z., Song, J., and Popper, A. N. (**1998**). "Encoding of acoustic directional information by saccular afferents of the sleeper goby, *Dormitator latifrons*," J. Comp. Physiol., A **182**, 805–815.

Lu, Z., and Xu, Z. (**2002**). "Effects of saccular otolith removal on hearing sensitivity of the sleeper goby (*Dormitator latifrons*)," J. Comp. Physiol., A **188**, 595–602.

Mann, D., Cott, P., Hanna, B. W., and Popper, A. N. (**2007**). "Hearing in eight species of northern Canadian freshwater fishes," J. Fish Biol. **70**, 109–120.

McCauley, R. D., Fewtrell, J., and Popper, A. N. (**2003**). "High intensity anthropogenic sound damages fish ears," J. Acoust. Soc. Am. **113**, 638–642.

Milne, A. R. (**1966**). "Statistical description of noise under shore-fast sea ice in winter," J. Acoust. Soc. Am. **39**, 1174–1182.

Popper, A., Smith, M., Cott, P., Hanna, B., MacGillivary, A., Austin, M., and Mann, D. (**2005**). "Effects of exposure to seismic airgun use on hearing of three fish species," J. Acoust. Soc. Am. **117**, 3958–3971.

Smith, M. E., Kane, A. S., and Popper, A. N. (**2004**). "Noise-induced stress response and hearing loss in goldfish (*Carassius auratus*)," J. Exp. Biol. **207**, 427–435.

Stein, P. J. (**1988**). "Interpretation of a few ice event transients," J. Acoust. Soc. Am. **83**, 617–622.

Stewart, D. B. (**2001**). "Possible impacts on overwintering fish of trucking granular materials over lake and river ice in the Mackenzie River delta area," prepared by Arctic Biological Consultants, Winnipeg, MB for Fisheries Joint Management Committee, Inuvik, NT.

Tavolga, W. N. (**1974**). "Signal/noise ratio and the critical band in fishes," J. Acoust. Soc. Am. **55**, 1323–1333.

Wysocki, L. E., and Ladich, F. (**2005**). "Hearing in fishes under noise conditions," J. Assoc. Res. Otolaryngol. **6**, 28–36.

Travel-time sensitivity kernels in long-range propagation

E. K. Skarsoulis
Institute of Applied and Computational Mathematics, Foundation for Research and Technology Hellas, P.O. Box 1385, 711 10 Heraklion, Crete, Greece

B. D. Cornuelle and M. A. Dzieciuch
Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0230

(Received 24 March 2009; revised 30 July 2009; accepted 19 August 2009)

Wave-theoretic travel-time sensitivity kernels (TSKs) are calculated in two-dimensional (2D) and three-dimensional (3D) environments and their behavior with increasing propagation range is studied and compared to that of ray-theoretic TSKs and corresponding Fresnel-volumes. The differences between the 2D and 3D TSKs average out when horizontal or cross-range marginals are considered, which indicates that they are not important in the case of range-independent sound-speed perturbations or perturbations of large scale compared to the lateral TSK extent. With increasing range, the wave-theoretic TSKs expand in the horizontal cross-range direction, their cross-range extent being comparable to that of the corresponding free-space Fresnel zone, whereas they remain bounded in the vertical. Vertical travel-time sensitivity kernels (VTSKs)—one-dimensional kernels describing the effect of horizontally uniform sound-speed changes on travel-times—are calculated analytically using a perturbation approach, and also numerically, as horizontal marginals of the corresponding TSKs. Good agreement between analytical and numerical VTSKs, as well as between 2D and 3D VTSKs, is found. As an alternative method to obtain wave-theoretic sensitivity kernels, the parabolic approximation is used; the resulting TSKs and VTSKs are in good agreement with normal-mode results. With increasing range, the wave-theoretic VTSKs approach the corresponding ray-theoretic sensitivity kernels.
 [DOI: 10.1121/1.3224835]

PACS number(s): 43.30.Pc, 43.30.Bp [JAC] Pages: 2223–2233

I. INTRODUCTION

The present work aims at studying the behavior of ocean acoustic travel-time sensitivity kernels (TSKs) in long-range propagation. The notion of TSKs was first introduced in the seismological literature[1,2] as a means to study the sensitivity of finite-frequency travel-time observables to changes in the sound-speed distribution,[3,4] as well as the relation between wave-theoretic and ray-theoretic travel-time observables,[5] taking into account that ray theory is a high-frequency asymptotic approximation. The TSK is associated with the first-order integral representation of travel-time variations in terms of the underlying spatial distribution of sound-speed variations, resulting from the first Born approximation of Green's function perturbations[6,7] and the notions of phase arrivals[1,2] or peak arrivals,[8,9] and represents the spatial distribution of sensitivity of arrival times to sound-speed variations.

The application of TSKs to short- and medium-range ocean acoustic propagation,[10] has revealed similarities to the sensitivity behavior of seismic travel times in global-scale propagation: small sensitivities in the immediate vicinity of the corresponding eigenray, negative sensitivities at a distance from the eigenray within the first Fresnel zone, followed by an alternating pattern of positive and negative sensitivities decaying with distance. This similarity between ocean acoustic and seismic TSKs at very different scales is dictated by the different intensity of refraction phenomena in the two cases. The vertical gradient of the sound speed in the water column (typically 0.017 s^{-1} at great depths[11]) is much larger than the corresponding gradient in the underlying earth (typically 0.001 s^{-1}).[4] This gives rise to stronger refraction and thus much shorter double-loop ranges in water (40–50 km) than in the solid earth (order of 1000 km).

Because of the smaller double-loop length characterizing ocean acoustic propagation, long-range acoustic paths sample the ocean much more uniformly in range and depth than seismic paths sample the solid earth underneath. Ocean acoustic tomography takes advantage of multipath propagation and exploits measured travel times, affected by the distribution of the sound speed, to retrieve the latter by inversion.[11,12] Travel times are integral measures of sound-speed changes along the corresponding propagation paths, with maximum sensitivity to changes around the corresponding turning depths (where the grazing angles are near zero).[10] In this connection, the resolving power of tomography applies mainly to the vertical structure of the ocean.[11]

The sensitivity of travel times to horizontally uniform changes in the vertical ocean structure is described through the vertical TSK (VTSK). The VTSK is a function of depth only and represents the first-order effect that a horizontally uniform change in the sound-speed profile at a particular depth has on finite-frequency travel times. The VTSKs can be obtained either analytically, by applying perturbation theory to the range-independent propagation problem and the associated Green's function,[10] or numerically, as marginals

of the two-dimensional (2D) and three-dimensional (3D) TSKs with respect to the horizontal. The aim of this work is to compare VTSKs calculated in different ways and to examine the TSK and VTSK behaviors as the propagation range increases. It is known, for example, that there are differences between the 2D and 3D TSKs, such as the zero-sensitivity cores of 3D TSKs along eigenrays, as opposed to the uniformly negative sensitivities of 2D TSKs.[2] Do these differences carry over to the case of the 2D and 3D VTSKs? Further, because ray theory is the most common approach for the interpretation of travel-time data from long-range propagation experiments, the question arises how wave-theoretic VTSKs compare to the corresponding ray-theoretic VTSKs at various propagation ranges; the latter are the same in two and three dimensions, and they are also frequency-independent (high-frequency asymptotic approximation).

The contents of this work are organized as follows. In Sec. II, the wave-theoretic modeling of travel times is reviewed and first-order expressions for their sensitivity kernels in two and three dimensions are given. Further, expressions for the VTSKs are derived by applying perturbation theory to the 2D and 3D Green's function normal-mode representations. Numerical results are presented in Sec. III demonstrating differences and similarities between 2D and 3D TSKs and VTSKs, as well as between VTSKs calculated from different modeling approaches [normal modes, parabolic approximation, and ray theory] or in different ways (analytically or numerically). The 3D TSKs are compared to Fresnel volumes, and the effects of increasing range on TSKs and VTSKs are addressed. In Sec. IV, the main results from this work are discussed and the basic conclusions are summarized.

II. DERIVATION OF TRAVEL-TIME SENSITIVITY KERNELS

In this section, the derivation of 2D and 3D TSKs based on the notion of peak arrivals and the first Born approximation is reviewed. Further, analytical expressions for the VTSKs are derived by applying perturbation theory to the normal-mode representation of the 2D and 3D Green's functions.

A. Wave-theoretic travel-time modeling and TSKs

Assuming a harmonic point source of circular frequency ω and unit strength at location $\vec{x}_s$ within an acoustic medium characterized by sound-speed distribution $c(\vec{x})$, the acoustic field at the location $\vec{x}_r$ (receiver location) is described by the Green's function $G(\vec{x}_r|\vec{x}_s;\omega;c)$. If the source emits a broadband signal $P_s(\omega)$, the corresponding acoustic complex pressure at the receiver in the time domain can be expressed through the inverse Fourier transform,

$$p_r(t;\vec{x}_s,\vec{x}_r;c)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}P_s(\omega)G(\vec{x}_r|\vec{x}_s;\omega;c)e^{i\omega t}d\omega. \quad (1)$$

A perturbation of the sound-speed distribution will result in a perturbation of the Green's function which to the first order is given by the first Born approximation.[6,7]

$$\Delta G_{3D}(\vec{x}_r|\vec{x}_s;\omega;c;\Delta c)=-2\omega^2\int\int_V\int G_{3D}(\vec{x}'|\vec{x}_s;\omega;c)\times G_{3D}(\vec{x}_r|\vec{x}';\omega;c)\frac{\Delta c(\vec{x}')}{c^3(\vec{x}')}dV(\vec{x}'). \quad (2)$$

This expression is valid in the 3D case. In the 2D case, the volume integral is replaced by an integral over the area where the sound-speed perturbation takes place

$$\Delta G_{2D}(\vec{x}_r|\vec{x}_s;\omega;c;\Delta c)=-2\omega^2\int\int_A G_{2D}(\vec{x}'|\vec{x}_s;\omega;c)\times G_{2D}(\vec{x}_r|\vec{x}';\omega;c)\frac{\Delta c(\vec{x}')}{c^3(\vec{x}')}dA(\vec{x}'). \quad (3)$$

The corresponding perturbation of the acoustic complex pressure p_r becomes

$$\Delta p_r(t;\vec{x}_s;\vec{x}_r;c;\Delta c)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}P_s(\omega)\Delta G(\vec{x}_r|\vec{x}_s;\omega;c;\Delta c)\times e^{i\omega t}d\omega. \quad (4)$$

For the wave-theoretic modeling of arrival times, the notion of peak arrivals[8,9] is used. Peak arrivals are defined as the significant local maxima of the arrival pattern (the amplitude of the complex pressure at the receiver in the time domain). Expressing the complex pressure in terms of its real and imaginary parts, $p_r=u+iw$, $\Delta p_r=\Delta u+i\Delta w$, the perturbation of peak arrival times can be expressed as[8]

$$\Delta\tau_\ell=-\frac{\dot{u}_\ell\Delta u_\ell+u_\ell\Delta\dot{u}_\ell+\dot{w}_\ell\Delta w_\ell+w_\ell\Delta\dot{w}_\ell}{\dot{u}_\ell^2+u_\ell\ddot{u}_\ell+\dot{w}_\ell^2+w_\ell\ddot{w}_\ell}, \quad (5)$$

where the index ℓ specifies the particular peak of interest in the received arrival pattern and the dots denote differentiation with respect to time.

Combination of Eqs. (5), (4), and (2) results in the following expression describing the first-order perturbations of travel times in terms of the underlying 3D sound-speed perturbations:[10]

$$\Delta\tau_\ell(\vec{x}_s,\vec{x}_r;c;\Delta c)=\int\int_V\int K_\ell(\vec{x}';\vec{x}_s,\vec{x}_r;c)\Delta c(\vec{x}')dV(\vec{x}'), \quad (6)$$

where the kernel $K_\ell(\vec{x}';\vec{x}_s;\vec{x}_r;c)$ is given by the expression

$$K_\ell(\vec{x}';\vec{x}_s,\vec{x}_r;c)=\frac{1}{b_\ell c^3(\vec{x}')}\Re\left\{(\dot{u}_\ell-i\dot{w}_\ell)\frac{1}{2\pi}\int_{-\infty}^{+\infty}2\omega^2P_s(\omega)\times G_{3D}(\vec{x}'|\vec{x}_s;\omega;c)G_{3D}(\vec{x}_r|\vec{x}';\omega;c)e^{i\omega\tau_\ell}d\omega+(u_\ell-iw_\ell)\frac{1}{2\pi}\int_{-\infty}^{+\infty}2i\omega^3P_s(\omega)G_{3D}(\vec{x}'|\vec{x}_s;\omega;c)\times G_{3D}(\vec{x}_r|\vec{x}';\omega;c)e^{i\omega\tau_\ell}d\omega\right\}, \quad (7)$$

with $b_\ell = \dot{u}_\ell^2 + u_\ell \ddot{u}_\ell + \dot{w}_\ell^2 + w_\ell \ddot{w}_\ell$. The kernel K_ℓ describes the first-order effect that a sound-speed perturbation at location $\vec{x}'$ has on the travel time τ_ℓ, and is known as the travel-time sensitivity kernel in three dimensions or 3D TSK. The 2D TSK is given by the same expression (7) in which G_{3D} is replaced by G_{2D}.

B. Normal-mode representation of the Green's function

Assuming a range-independent environment in which the speed of sound is a function of depth z only, $c=c(z)$, the Green's function for the far field can be expressed in terms of the real eigenvalues k_m and corresponding eigenfunctions $\varphi_m(z)$ of the vertical Sturm–Liouville problem,[13,14]

$$\frac{d^2\varphi_m(z)}{dz^2} + \frac{\omega^2}{c^2(z)}\varphi_m(z) = k_m^2 \varphi_m(z), \tag{8}$$

supplemented by the conditions that $\varphi_m = 0$ at the sea surface $(z=0)$, φ_m and $\rho^{-1} d\varphi_m/dz$ are continuous across the interfaces, and φ_m and $d\varphi_m/dz$ are vanishing as $z \to \infty$, where ρ is the density and ω is the circular frequency of the source.

1. Three-dimensional propagation problem

Using a cylindrical coordinate system (r, z, ϑ) with origin at the sea surface, the vertical z-axis positive downwards, and the source on this axis at depth z_s with time dependence $e^{i\omega t}$, see Eq. (1), the far-field Green's function in the water at location (r,z) can be written in the form[14]

$$G_{3D}(r,z|z_s) = \frac{e^{-i\pi/4}}{\rho_W \sqrt{8\pi}} \sum_{m=1}^{M} \frac{\varphi_m(z_s)\varphi_m(z)}{\sqrt{k_m r}} e^{-ik_m r}, \tag{9}$$

where ρ_W is the water density. The difference in the signs of the exponential arguments between Eq. (9) and Ref. 14 is due to the different time dependences ($e^{-i\omega t}$ in Ref. 14). Substitution of Eq. (9) into Eq. (7) gives the 3D TSK which describes to the first order the effect that a unit sound-speed perturbation over a unit volume about location $\vec{x}'$ has on travel times.

2. Two-dimensional propagation problem

Using a Cartesian coordinate system (x,z) with origin at the sea surface, the vertical z-axis positive downwards, and the source on this axis at depth z_s with time dependence $e^{i\omega t}$, the far-field Green's function in the water at location (x,z) can be written as[14]

$$G_{2D}(x,z|z_s) = \frac{e^{-i\pi/2}}{2\rho_W} \sum_{m=1}^{M} \frac{\varphi_m(z_s)\varphi_m(z)}{k_m} e^{-ik_m x}. \tag{10}$$

Concerning the signs of the exponentials, the same comment applies as before. The resulting TSK in the two-dimensional problem is a function of range and depth and applies to sound-speed perturbations distributed over 2D areas in the range-depth plane.

C. Green's function from the parabolic approximation

The parabolic approximation enables the efficient calculation of the Green's function (and the TSK) in range-dependent environments. Neglecting the azimuthal dependence of the ocean environment and the acoustic field, and adopting a Cartesian coordinate system (x,z), the parabolic equation (PE) for the acoustic pressure associated with the outgoing energy in two dimensions is[15]

$$\frac{\partial p}{\partial x} = ik_0 \sqrt{1+X}\, p, \tag{11}$$

where X is a differential operator with respect to the depth variable

$$X = \frac{1}{k_0^2}\left(\frac{\partial^2}{\partial z^2} + k^2 - k_0^2\right), \tag{12}$$

$k = \omega/c(x,z)$, and $k_0 = \omega/c_0$, where c_0 is a representative sound-speed value.

For the evaluation of the 2D Green's function the RAM code[15,16] is used. This code solves Eq. (11) by applying a rational approximation to the exponential of the square-root operator $\exp(\sqrt{1+X})$ and using a marching scheme for integration in range. Furthermore, it uses a starting field which allows for wide propagation angles;[17] this is particularly important for the accurate evaluation of the phase and travel-times associated with acoustic energy propagating at large grazing angles.

D. Vertical travel-time sensitivity kernel

In the case of a range-independent problem, in which both the background and the perturbed state are range independent, the VTSK describes the first-order effect that a perturbation of the sound-speed profile at depth z' will have on the corresponding travel time:

$$\Delta\tau_\ell(\vec{x}_s, \vec{x}_r; c; \Delta c) = \int_0^H D_\ell(z'; \vec{x}_s, \vec{x}_r; c)\Delta c(z')dz', \tag{13}$$

where H is the water depth and $D_\ell(z'; \vec{x}_s, \vec{x}_r; c)$ is the first-order VTSK. Since both the background and the perturbed environment are range-independent, the Green's function for the perturbed environment can be expressed in terms of normal modes, as in Sec. II B. Applying perturbation theory[18] to the depth problem, the following expressions for the first-order perturbations of eigenvalues and eigenfunctions can be obtained:

$$\Delta k_{m,1} = \frac{Q_{mm}}{2k_m}, \tag{14}$$

$$\Delta\varphi_{m,1}(z) = \sum_{\substack{n=1 \\ n\neq m}}^{M} \frac{Q_{mn}\varphi_n(z)}{k_m^2 - k_n^2}, \tag{15}$$

where

$$Q_{mn} = -2\omega^2 \int_0^h \frac{\varphi_m(z')\varphi_n(z')}{c^3(z')}\Delta c(z')dz'. \tag{16}$$

1. Three-dimensional propagation problem

Taking the first-order perturbation of the three-dimensional Green's function [Eq. (9)] and substituting the above perturbation expressions for the eigenvalues and eigenfunctions, the following perturbation formula can be obtained:

$$\Delta G_{3D} = \frac{e^{-i\pi/4}}{\rho_W \sqrt{8\pi}} \sum_{m=1}^{M} \left[\sum_{\substack{n=1 \\ n\neq m}}^{M} \frac{Q_{mn} U_{mn}}{\Lambda_{mn}} + \left(\frac{1}{2k_m} + ir \right) \frac{U_{mm} Q_{mm}}{k_m} \right] \frac{e^{-ik_m r}}{\sqrt{k_m r}}, \tag{17}$$

where $\Lambda_{mn} = k_m^2 - k_n^2$ and

$$U_{mn} = \begin{cases} \varphi_m(z_S)\varphi_n(z) + \varphi_n(z_S)\varphi_m(z), & n \neq m \\ -\frac{1}{2}\varphi_m(z_S)\varphi_m(z), & n = m. \end{cases} \tag{18}$$

The dominating term in Eq. (17) for propagation over long ranges is the one involving ir in the parentheses. Substituting the above expression into Eq. (4), and exploiting the perturbation relation (5) for peak arrivals, the VTSK for propagation in a three-dimensional range-independent environment takes the form

$$D_\ell^{(3D)}(z';\vec{x}_s,\vec{x}_r;c) = \frac{e^{-i\pi/4}}{\rho_W b_\ell c^3(z')\sqrt{2\pi r}} \Re\left\{ (\dot{u}_\ell - i\dot{w}_\ell) \frac{1}{2\pi} \int_{-\infty}^{+\infty} L^{(3D)}(z';\omega) \times e^{i\omega\tau_\ell} d\omega + (u_\ell - iw_\ell) \frac{1}{2\pi} \int_{-\infty}^{+\infty} i\omega L^{(3D)}(z';\omega)^{i\omega\tau_\ell} d\omega \right\}, \tag{19}$$

where

$$L^{(3D)}(z';\omega) = \omega^2 P_S(\omega) \sum_{m=1}^{M} \left[\sum_{\substack{n=1 \\ n\neq m}}^{M} \frac{\varphi_m(z')\varphi_n(z')U_{mn}}{\Lambda_{mn}} + \left(\frac{1}{2k_m} + ir \right) \frac{U_{mm}\varphi_m^2(z')}{k_m} \right] \frac{e^{-ik_m r}}{\sqrt{k_m r}}. \tag{20}$$

The 3D VTSK can be alternatively obtained as the marginal of the 3D TSK with respect to the horizontal (integration with respect to range and azimuth). Comparisons between the analytic 3D VTSK, Eq. (19), and the horizontal marginal of the 3D TSK are made in Sec. III.

2. Two-dimensional propagation problem

Taking the first-order perturbation of the two-dimensional Green's function [Eq. (10)] and substituting the perturbation expressions (14) and (15) for the eigenvalues and eigenfunctions, the following perturbation formula can be obtained:

$$\Delta G_{2D} = \frac{-i}{2\rho_W} \sum_{m=1}^{M} \left[\sum_{\substack{n=1 \\ n\neq m}}^{M} \frac{Q_{mn} U_{mn}}{\Lambda_{mn}} + \left(\frac{1}{k_m} + ix \right) \frac{U_{mm} Q_{mm}}{k_m} \right] \frac{e^{-ik_m x}}{k_m}. \tag{21}$$

The dominating term for propagation over long ranges is the one involving ix in the parentheses. Substituting the above expression into Eq. (4) and exploiting the perturbation relation (5) for peak arrivals the VTSK for propagation in a two-dimensional range-independent environment takes the form

$$D_\ell^{(2D)}(z';\vec{x}_s,\vec{x}_r;c) = \frac{-i}{\rho_W b_\ell c^3(z')} \Re\left\{ (\dot{u}_\ell - i\dot{w}_\ell) \frac{1}{2\pi} \int_{-\infty}^{+\infty} L^{(2D)}(z';\omega) \times e^{i\omega\tau_\ell} d\omega + (u_\ell - iw_\ell) \frac{1}{2\pi} \int_{-\infty}^{+\infty} i\omega L^{(2D)}(z';\omega)^{i\omega\tau_\ell} d\omega \right\}, \tag{22}$$

where

$$L^{(2D)}(z';\omega) = \omega^2 P_S(\omega) \sum_{m=1}^{M} \left[\sum_{\substack{n=1 \\ n\neq m}}^{M} \frac{\varphi_m(z')\varphi_n(z')U_{mn}}{\Lambda_{mn}} + \left(\frac{1}{k_m} + ix \right) \frac{U_{mm}\varphi_m^2(z')}{k_m} \right] \frac{e^{-ik_m x}}{k_m}. \tag{23}$$

In Sec. III, the analytic 2D VTSK, Eq. (22), is compared with the marginal of the 2D TSK with respect to the horizontal, as well as with the analytic 3D VTSK, Eq. (19).

III. NUMERICAL RESULTS

Some numerical results are presented in this section to shed light on the differences and similarities between the various TSKs and VTSKs, and to show the effects of increasing range.

For their presentation, the wave-theoretic TSKs are calculated on a rectangular grid and they are smoothed to suppress small-scale oscillations. In the vertical and horizontal cross-range directions, the grid spacing is 10 m and the smoothing window is 100 m. In the horizontal along-range direction, the grid spacing is 200 m and the smoothing window is 600 m, unless otherwise mentioned. The same parameters are used for the VTSKs (spacing in depth of 10 m and smoothing window of 100 m).

A. Comparison between the different kernels

For the comparison between the different kernels, a simple environment is considered characterized by a linear sound-speed profile, 1503 m/s at the surface and 1547 m/s at the depth of 2500 m, typical of western Mediterranean propagation conditions in winter, followed by an absorbing

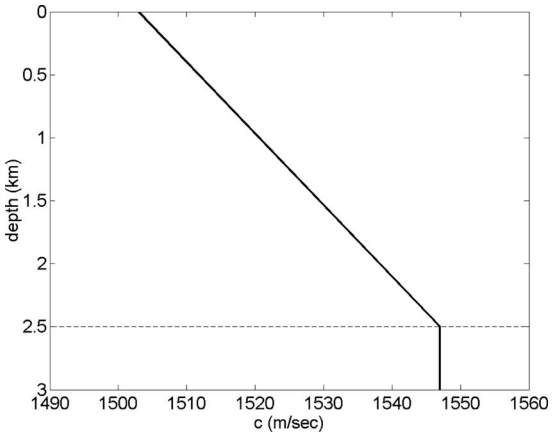


FIG. 1. Linear sound-speed profile—water depth 2500 m—and absorbing bottom.

bottom (half space of unit density and constant sound speed, equal to the water sound speed at the water-bottom interface), see Fig. 1. The signal emitted by the source is a Gaussian pulse of 100-Hz central frequency and 50-Hz bandwidth (3-dB bandwidth). Both source and receiver are considered at a depth of 150 m and at various ranges.

Figure 2 shows the predicted 2D and 3D arrival patterns for source/receiver range of 100 km (top panels) and the corresponding TSKs on the source-receiver vertical plane (lower panels) for the first three arrivals based on normal-mode calculations. The results on the left are based on the 3D Green's function, whereas the ones on the right are from the 2D Green's function. The Green's functions have been evaluated at 1901 frequencies, from 5 to 195 Hz with step 0.1 Hz—the calculation bandwidth has been taken broad enough to avoid clipping of significant parts of the signal spectrum (and thus avoid side lobes in the time domain). The wave-theoretic TSKs are shown in color, whereas the black dashed lines represent the corresponding eigenrays. The first and third peaks are simple arrivals corresponding to single eigenrays connecting the source and the receiver, whereas the second peak is a double arrival corresponding to two symmetric eigenrays. The color scales range from blue (negative sensitivity) to red (positive sensitivity) with green corresponding to zero sensitivity. Negative sensitivity means that a sound-speed increase will lead to a travel-time decrease—this is the anticipated behavior. Positive sensitivity, on the other hand, indicates that a sound-speed increase will lead to a travel-time increase. Even though the areas of negative sensitivity are prevailing in the wave-theoretic TSKs, there are areas of positive sensitivity as well.[10]

While the predicted 2D and 3D arrival patterns in Fig. 2 are very similar, the corresponding 2D and 3D TSKs are quite different. Close to the turning points, the 3D TSK is near zero on the eigenrays and takes negative values at a distance, whereas the 2D TSK is uniformly negative. This difference between 3D and 2D TSKs was first observed in the modeling of seismic travel-time observables (phase arrivals) and suggests that sound-speed changes taking place on the eigenrays do affect travel times in two dimensions but have no effect in three dimensions—this is known as the banana-doughnut paradox.[2] Further, from Fig. 2 it is seen that the negative-sensitivity domain of the 3D TSK is thicker than that of the 2D TSK in the neighborhood of the turning points but thinner elsewhere. The different sensitivity magnitudes in the 3D and 2D cases (different color scales) have to do with the fact that the 3D TSKs apply to sound-speed perturbations over 3D volumes, whereas the 2D TSKs apply to perturbations over 2D areas.

The 3D Green's function describes the acoustic field of a point source in 3D space, whereas the 2D Green's function

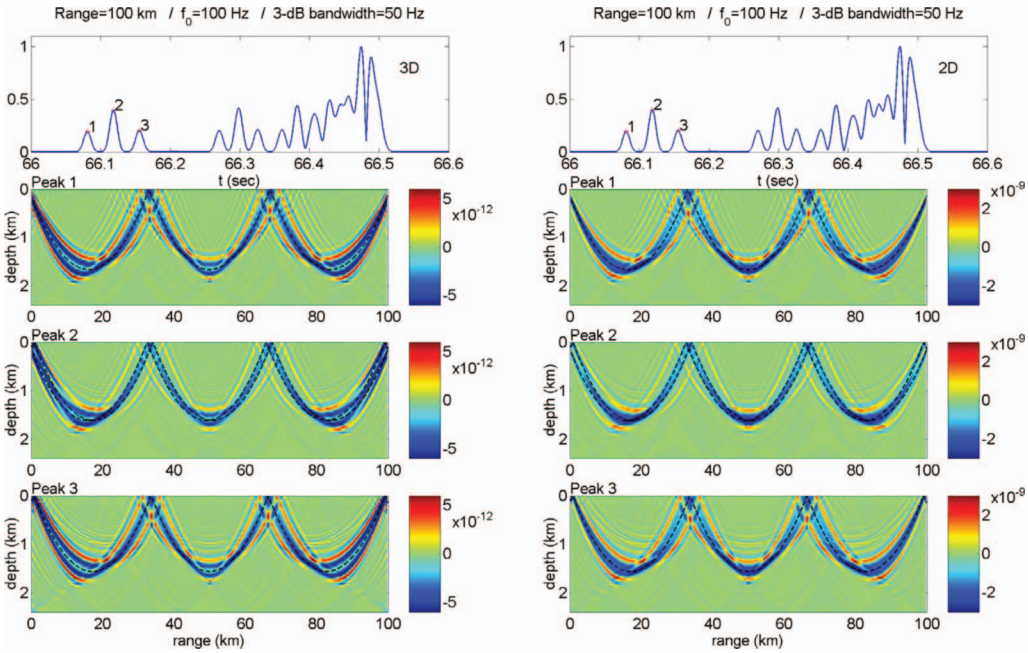


FIG. 2. Wave-theoretic arrival pattern for source-receiver range of 100 km (top) and corresponding TSKs for the first three peak arrivals (lower panels) for linear sound-speed profile, based on normal-mode calculations in the 3D (left) and 2D (right) cases. The units of the 2D and 3D TSKs are s^2/m^3 and s^2/m^4, respectively.

describes the field of a line source in 3D space, in which case the field on any slice normal to the line is the same. In this connection, Fig. 2 describes the 2D TSK completely but shows only a single section of the 3D TSK on the vertical plane containing the source and the receiver. The 3D TSK is symmetric about this plane and changes with distance from it. Thus, based on Fig. 2 alone, one cannot answer the question how significant the above-mentioned differences between the 2D and 3D TSKs are when applied to sound-speed perturbations of scales that are large compared to the lateral extent of the TSKs. To answer that question, the marginals of the 2D and 3D TSKs in the horizontal and cross-range directions are evaluated in the following.

Figure 3(a) shows the cross-range marginals of the 2D and 3D TSKs corresponding to Peak 2 of Fig. 2. Cross-range marginal means integration in the vertical for the 2D TSK, whereas for the 3D TSK it means integration in the vertical and azimuthal directions. These integrations are carried out numerically with discretization step 10 m in the vertical and azimuthal directions, within ± 4 km from the central vertical plane (for the 3D TSK). The azimuthal integration takes into account the axial symmetry of the Green's functions in Eq. (2) near the source and receiver location, whereas in the far field it coincides with integration perpendicular to the central vertical plane. The result is a horizontal travel-time sensitivity kernel (HTSK) which is a function of range. The two HTSKs shown in Fig. 3(a), resulting as marginals of the 2D and 3D TSKs, are in good agreement at all ranges, which means that the differences between 2D and 3D TSKs do not carry over to their cross-range marginals. Away from the source and the receiver, the HTSKs are subject to very small variability, between -4.28×10^{-7} and -4.51×10^{-7} $\mathrm{s}^2/\mathrm{m}^2$, with slightly lower sensitivity at the ranges of the lower turning points. The limited variability of the HTSK with range points to very small sensitivity of travel times to the range position of sound-speed perturbations, in other words lack of horizontal resolution. For comparison, the ray-theoretic sensitivity $-1/c_0^2$ for a homogeneous medium with $c_0 = 1510$ m/s is also presented in Fig. 3(a) (axial line). The wave-theoretic HTSKs are concentrated about that line.

Figure 3(b) shows the horizontal marginals of the 2D and 3D TSKs corresponding to Peak 2 of Fig. 2. Horizontal marginal means integration over range and azimuth for the 3D TSK and integration over range for the 2D TSK. The result is a function of depth which describes the effect that a horizontally uniform sound-speed perturbation at any particular depth has on travel times (VTSK). The 2D and 3D marginals in Fig. 3(b) are in good agreement at all depths, despite the differences between the 2D and 3D TSKs. The VTSK attains its maximum negative value close to the corresponding ray-theoretic lower turning depth (1580 m); i.e., the travel time is more sensitive to sound-speed changes around the corresponding turning depth. While negative sensitivity prevails, there are intervals, below the turning depth, where positive sensitivity is observed.

The agreement between the horizontal and cross-range marginals of the 2D and 3D TSKs indicates that the differences between the 2D and 3D TSKs shown in Fig. 2 average out in the case of range-independent perturbations or perturbations with scales that are large compared to the TSK lateral extent. In this connection, the wave-theoretic treatment of travel-time observables using a 2D or 3D approach is not expected to be significantly different. This also brings the wave-theoretic VTSKs a step closer to the ray-theoretic behavior, in which there are no differences between two and three dimensions.

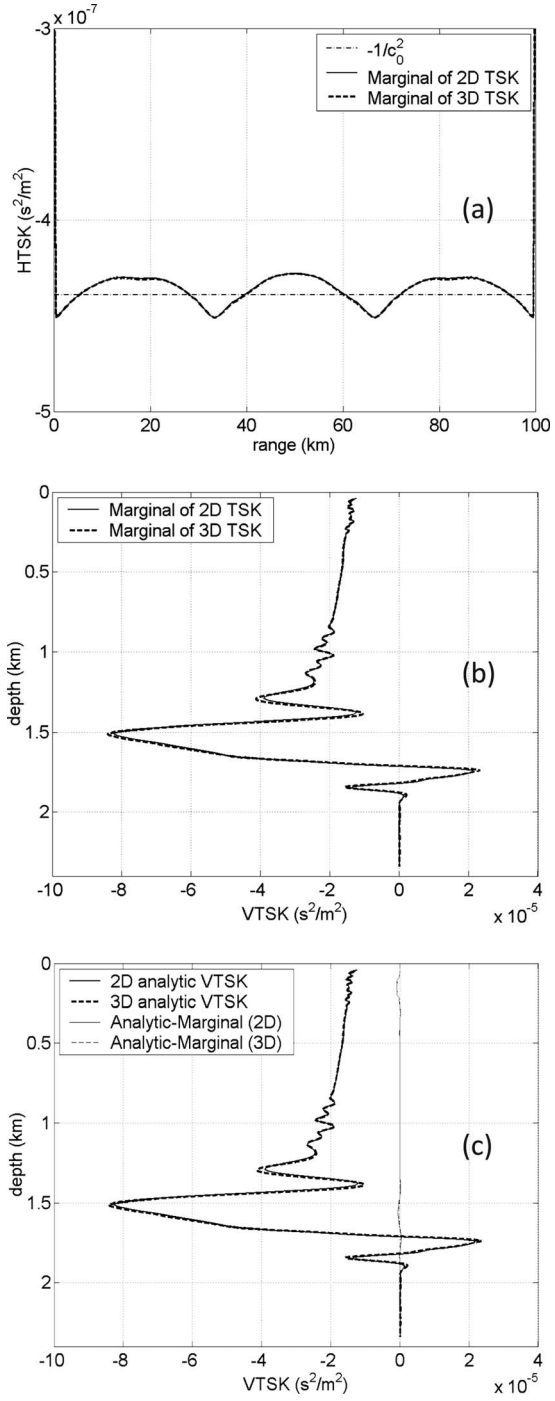


FIG. 3. (a) Cross-range marginals of the 2D and 3D TSKs for Peak 2 (Fig. 2) and theoretic sensitivity $(-1/c_0^2)$ for a homogeneous medium with $c_0 = 1510$ m/s. (b) Horizontal marginals of the 2D and 3D TSKs. (c) Analytic 2D and 3D VTSKs based on normal-mode calculations; the light lines show the differences from the VTSKs of the middle panel.

Figure 3(c) shows the 2D and 3D VTSKs from the analytical calculation, based on perturbation of the 2D and 3D

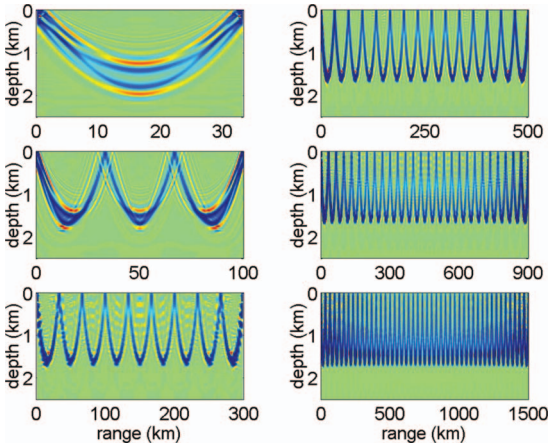


FIG. 4. Vertical along-range sections of 3D TSK of early arrivals with identical turning depth (1580 m) for linear sound-speed profile and for six propagation ranges between 33.3 and 1500 km.

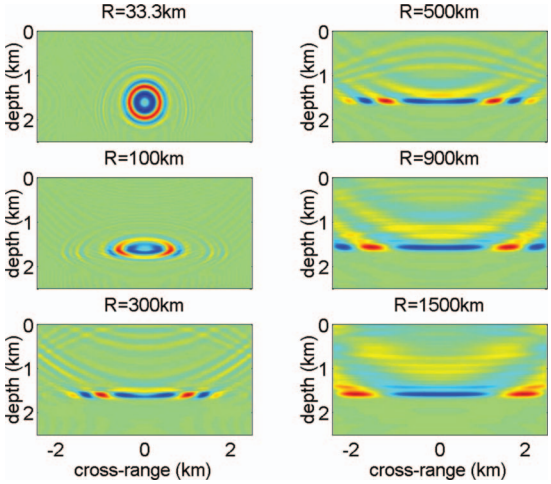


FIG. 5. Vertical cross-range sections of the 3D TSKs shown in Fig. 4 at mid-distance between source and receiver.

Green's function (Sec. II D). The calculation refers to the same environment, propagation range, and peak as before. The agreement between the analytic 2D and 3D VTSKs is very good, indicating the equivalence between the 2D and 3D wave-theoretic approaches for the retrieval of range-independent perturbations in the sound-speed profile from travel-time data. The light lines in Fig. 3(c) show the differences between the analytical 2D and 3D VTSKs and those obtained as horizontal marginals of the 2D and 3D TSKs [Fig. 3(b)]. The differences are very small and confirm the equivalence of the two ways of calculation (analytical and numerical) of the VTSKs.

B. Effects of increasing range

In this subsection, the effects of increasing range on TSKs and VTSKs are addressed. Figure 4 shows vertical sections (in the source/receiver vertical plane) of the wave-theoretic 3D TSK corresponding to early arrivals for six different source-receiver separations (propagation ranges): 33.3 (100/3), 100, 300, 500, 900, and 1500 km. To keep the burden for the long-range computations within limits, for this particular figure a horizontal resolution and smoothing window of 0.1% and 0.3% of the propagation range, respectively, were used. The color scale ranges from blue (negative sensitivity) to red (positive sensitivity), with green corresponding to zero sensitivity, and spans the variability range of each subplot. The arrivals are selected such that they all correspond to the same turning depth, that of the second arrival in the 100-km case, see Fig. 2 (surface-reflected arrival corresponding to two eigenrays with three complete double loops and lower turning depth of 1580 m). By taking integral multiples and submultiples of the 100-km range, the resulting arrivals correspond to eigenrays with the same launch angles, i.e., with the same turning depths. For example, the arrival for the range of 100/3 km corresponds to one double loop, the arrival for 300 km to nine double loops, etc., all with the same turning depth (1580 m).

For the shorter range of 100/3 km, the typical banana-doughnut picture is reproduced with a zero-sensitivity core (with maximum extent about the turning point) surrounded by a domain of negative sensitivity which is followed by positive sensitivity further out. At a range of 100 km, the TSK contracts in the vertical with simultaneous significant reduction in the zero-sensitivity cores. For longer ranges, the zero-sensitivity cores disappear leaving behind an area of near uniform negative sensitivity and a picture close to ray-theoretic. This behavior of the sensitivity kernel is quite different from the free-space behavior—in free space the zero-sensitivity core about the source-receiver connecting line persists and its cross-range extent increases proportionally to the square root of the source-receiver range.[10]

Figure 5 shows vertical cross-range sections of the travel-time sensitivity kernels shown in Fig. 4 at midrange between source and receiver, corresponding to lower turning points. This figure illustrates how the cross-section of the TSK behaves with increasing propagation range. At 100/3 km, the sensitivity kernel cross-section at midrange has a nearly circular symmetry with near-zero sensitivity in the middle. As the source-receiver range increases, the TSK contracts in the vertical and expands in the horizontal, passing from circular symmetry to ellipse-shaped and finally linear cross-section, whereas the near-zero sensitivity area in the middle gradually disappears.

Figure 6 shows vertical sections of the 3D TSK for the same propagation ranges as before, but now focusing on the first 100 km of propagation, presenting in detail the extent of the sensitivity areas about the eigenrays (marked by the dashed black lines). The solid black lines represent the boundaries of the Fresnel volumes[19] for frequency of 100 Hz. In a recent paper, Rypina and Brown[20] calculated Fresnel volumes using a second-order expansion of travel times with respect to longitudinal perturbations of the scattering location. For the calculations performed here, the exact dependence of travel times on lateral perturbations of the scattering location is evaluated for each location along the eigenrays; this dependence turns out to be strongly non-linear and multi-valued near the turning points. The Fresnel-volume boundaries shown in Fig. 6 correspond to phase difference from the central eigenrays equal to π.

For the range of 100/3 km, the boundary of the Fresnel

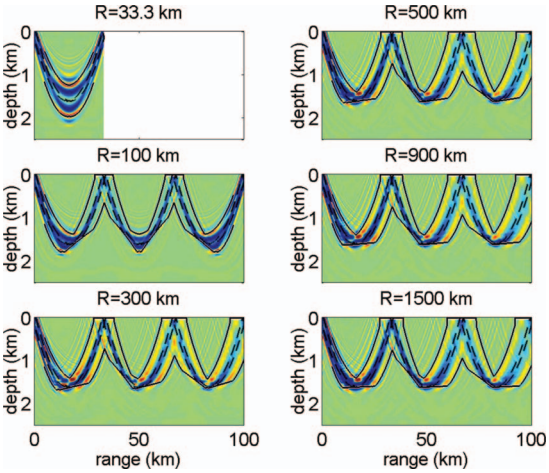


FIG. 6. 3D TSKs for linear sound-speed profile and for six propagation ranges (as in Fig. 4) focusing on the first 100 km of propagation. The dashed lines represent the corresponding eigenrays, and the solid lines represent the Fresnel-volume boundaries.

volume coincides with the positive sensitivity area of the TSK in agreement with the anticipated behavior in free space.[10] As the range increases to 100 km, the Fresnel volume contracts in the vertical toward the eigenray, similar to the 3D TSK, while its lower boundary is constrained by the presence of caustics, especially close to the turning points. A similar behavior can be observed in the wave-theoretic TSK near the turning points as well. As the source-receiver range increases to 300 km, the zero-sensitivity cores disappear leaving behind domains of uniformly negative sensitivity (blue color) surrounded by domains of positive sensitivity (yellow and red colors), which now fall within the boundaries of the Fresnel volumes. For longer ranges, the TSK and Fresnel volume characteristics in the first 100 km of propagation remain nearly unchanged.

Figure 7 focuses on a 100-km window about the midrange of propagation between source and receiver. Both the wave-theoretic TSKs and the Fresnel volumes appear to expand with increasing range, still their vertical extent is

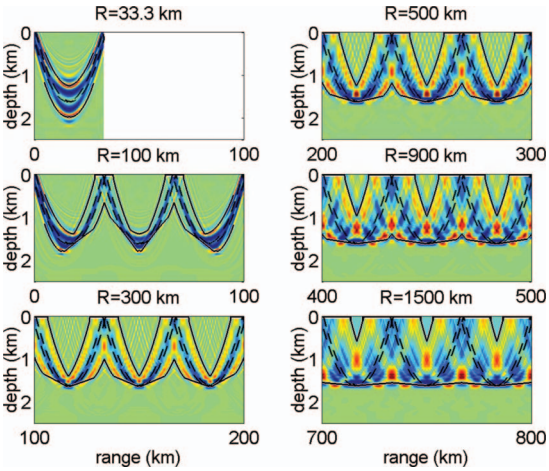


FIG. 7. 3D TSKs for linear sound-speed profile and for six propagation ranges (as in Fig. 4) focusing about the middle of source-receiver distance. The dashed lines represent the corresponding eigenrays, and the solid lines represent the Fresnel-volume boundaries.

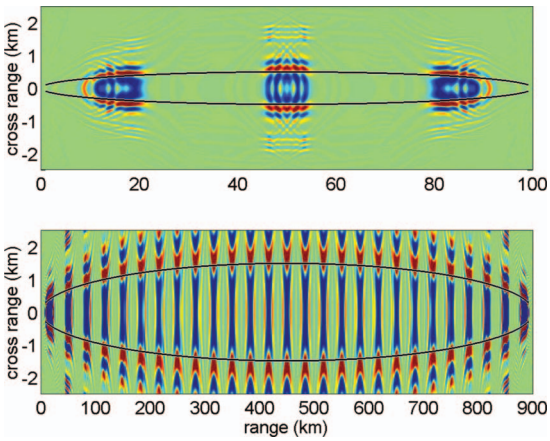


FIG. 8. Horizontal section (top view) of 3D TSKs for linear sound-speed profile and for source-receiver ranges of 100 and 900 km at the depth of 1580 m (turning depth). The solid lines represent the first Fresnel zone assuming free-space propagation between source and receiver.

limited because of refraction and the existence of caustics. Thus, while both the 3D TSK and the Fresnel volume expand down to depths of 2000 m for the short propagation range of 100/3 km, in the case of long-range propagation, they hardly reach depths beyond the ray-theoretic turning depth (1580 m). On the other hand, at the turning points and for the longer ranges, the negative-sensitivity area of the wave-theoretic TSKs is closely concentrated about the ray-theoretic turning depth (see also Fig. 5), whereas the corresponding Fresnel volumes have a much larger vertical extent toward the surface. A behavior similar to that of the wave-theoretic TSKs near the turning points in Fig. 7 is observed in the TSKs calculated from the parabolic approximation in Ref. 20 for propagation range 990 km (for small and intermediate values of the stability parameter α).

Figure 8 shows a horizontal section (top view) of the 3D TSKs at the turning depth (1580 m) for propagation ranges of 100 and 900 km. As in the case of Fig. 4, to limit the computational burden a range resolution and corresponding smoothing window 0.1% and 0.3% of the propagation range, respectively, were used. The traces of the acoustic paths at the turning depth can be seen in this figure, with the negative-sensitivity cores (in blue) surrounded by positive-sensitivity areas (in red). The cross-range extent of the negative TSK cores forms an elliptical envelope, which falls within the corresponding free-space Fresnel zone[10] (marked by the black solid lines) indicating that the horizontal cross-range behavior/extent of the travel-time sensitivity kernel at the turning depth is not affected by stratification. This is quite expected because the range-independent background environment under study does not give rise to horizontal refraction. Further, the Fresnel boundary intersects with the positive-sensitivity areas, as in the case of free-space TSKs.[10] Figures 5–8 indicate that at the turning points the horizontal cross-range extent of the negative-sensitivity core of the wave-theoretic TSKs prevails over the corresponding vertical extent.

In the following, some results for the VTSKs are presented, focusing on the effect of increasing range and the comparison between different modeling approaches. Figure 9

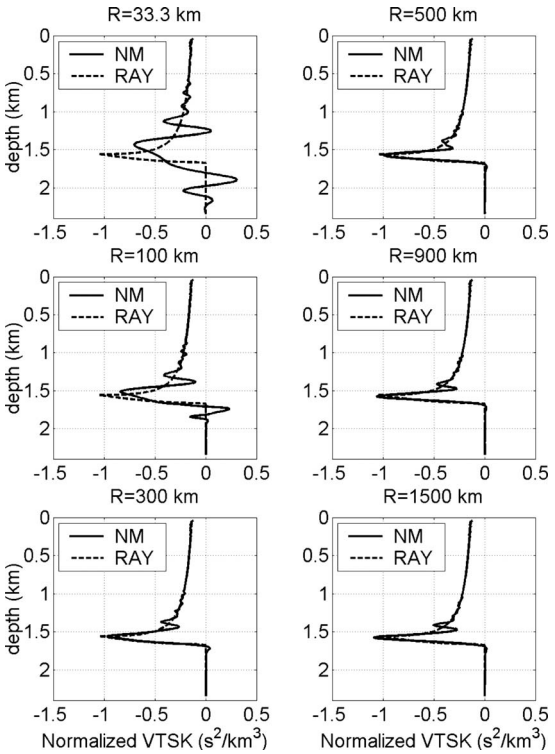


FIG. 9. Normalized VTSKs of early arrivals with identical turning depth (1580 m) for linear sound-speed profile and for six propagation ranges (as in Fig. 4) from normal-mode and ray-theoretic calculations—smoothed. Normalization consists in division with propagation range.

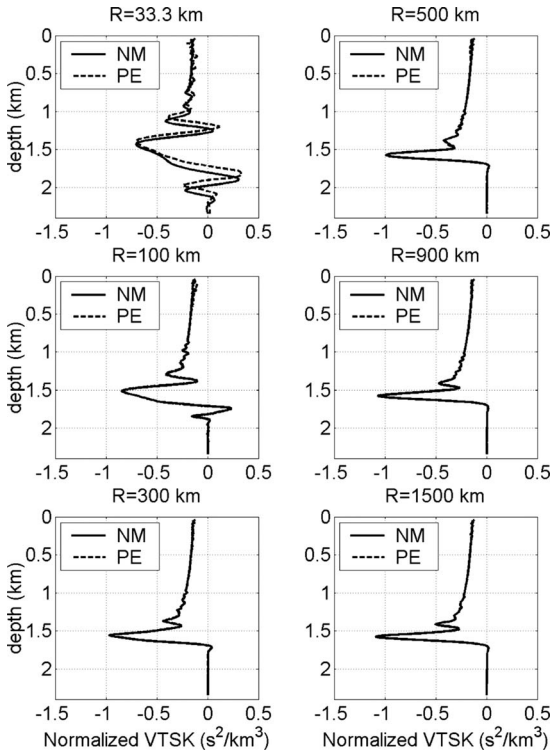


FIG. 10. Normalized VTSKs of early arrivals with identical turning depth (1580 m) for linear sound-speed profile and for six propagation ranges (as in Fig. 4) from normal-mode and PE calculations—smoothed. Normalization consists in division with range.

shows a comparison of normalized VTSKs from normal-mode and ray-theoretic calculations for source-receiver ranges of 33.3 (100/3), 100, 300, 500, 900, and 1500 km and for the same arrivals as in Figs. 4–7. Normalization here means division with the propagation range. The ray-theoretic VTSK is calculated as the horizontal marginal of the corresponding TSK, which equals $-1/c^2(z)$ along the eigenray (resulting from travel-time perturbations of ray arrivals).[21] Both normal-mode and ray-theoretic VTSKs are smoothed using the same 100-m moving average in depth (due to this the ray-theoretic VTSKs are not exactly horizontal at the turning depth).

For the short propagation range (100/3 km), the wave-theoretic VTSK exhibits significant deviations from the ray-theoretic one, both in amplitude and in the vertical extent: While the ray-theoretic sensitivity is strictly negative from the surface down to the turning depth, corresponding to the physical expectation that a sound-speed increase should cause a travel-time decrease, and zero elsewhere, the wave-theoretic VTSK, though mainly negative, takes also positive values over certain depth intervals. Thus, in the wave-theoretic context, a sound-speed increase may also cause a travel-time increase if it takes place around a depth of positive sensitivity. Further, the wave-theoretic TSK extends by more than 500 m below the ray-theoretic turning depth, which means that finite-frequency travel times are sensitive to sound-speed changes taking place at depths far beyond the ray-theoretic turning depth.

Coming to the effect of increasing range, from Fig. 9, it is seen that as the propagation range increases the wave-theoretic normalized VTSK approaches the ray-theoretic normalized VTSK; note that the latter remains the same at all ranges. For ranges beyond 500 km, the two sensitivity kernels are very close to each other, and this is quite remarkable taking into account that ray theory is a high-frequency asymptotic approximation while the frequency here is low (100 Hz). This suggests that for long-range propagation, ray-theoretic VTSKs can be applied even in the case of low frequencies.

Figure 10 shows a comparison of normalized VTSKs from normal-mode and parabolic approximation (PE) calculations. The PE VTSKs are obtained through the horizontal marginals of the corresponding 2D TSKs (not shown here). It is seen that the normal-mode and PE results (both wave-theoretic) are very close to each other in all cases—the agreement is similar in the case of the 2D TSKs—and only for the shortest range (100/3 km) there is a small deviation in the form of a phase shift. This deviation is attributed to the different starting fields of the normal-mode and PE solutions. The PE starting field is the sum of the discrete normal-mode field plus the near-field (or continuous) steeper-angle field; this field decays with range resulting in a good agreement for longer ranges. This agreement gives credibility to the wave-theoretic results and further supports the conclusion on the sufficiency of ray-theoretic VTSKs for long-range propagation even at low frequencies.

The results presented up to this point refer to a linear sound-speed profile. Some results for a temperate profile are presented in the following. The sound-speed profile, shown in Fig. 11, is typical for the north Pacific Ocean and has an

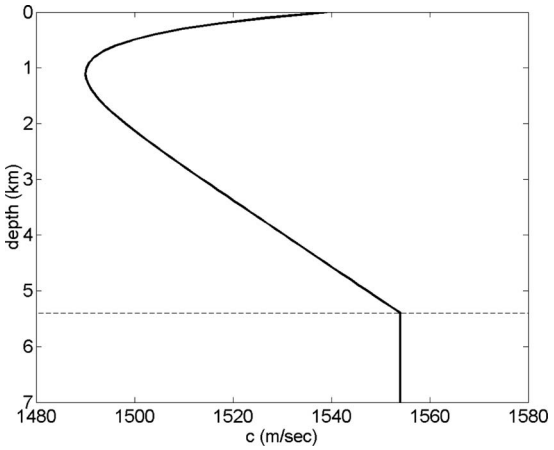


FIG. 11. Temperate sound-speed profile—water depth 5400 m—and absorbing bottom.

axis at 1100-m depth. The water depth in this case is 5400 m, and below that depth an absorbing bottom is considered as in the previous case. The source and receiver are taken both at axial depth (1100 m). The signal characteristics are the same as in the previous case. Figure 12 shows normalized VTSKs from normal-mode and ray-theoretic calculations for source-receiver ranges of 50, 100, 300, 500, 900, and 1500 km for refracted arrivals with the same upper and lower turning depths, 315 and 2704 m, respectively. These are double arrivals with double-loop range equal to 50 km, so the above propagation ranges correspond to 1, 2, 6, 10, 18, and 30 double loops, respectively.

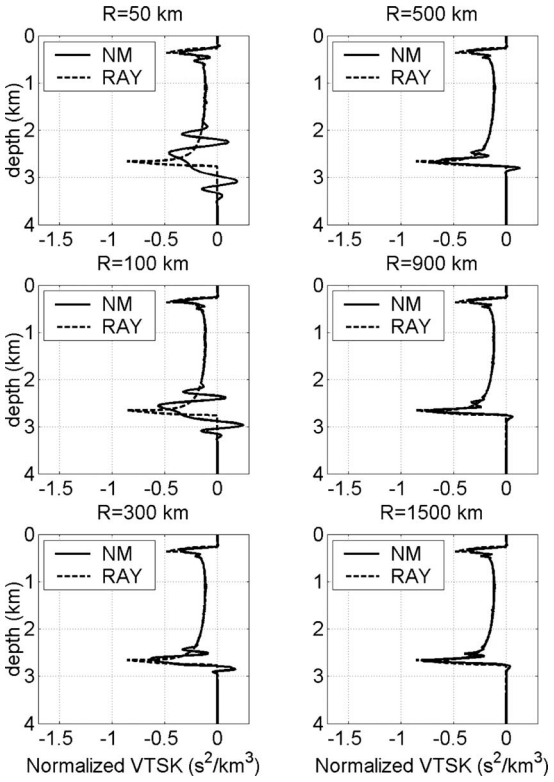


FIG. 12. Normalized VTSKs of early arrivals with identical upper and lower turning depths (315 and 2704 m) for temperate sound-speed profile and for six propagation ranges from normal-mode and ray-theoretic calculations—smoothed. Normalization consists in division with propagation range.

The ray-theoretic normalized VTSKs are the same for all ranges. They are supported between the turning depths, where they reach maximum negative values, whereas their minimum sensitivity is at the axial depth where the ray steepness with respect to the horizontal is largest. The sensitivity at the lower turning depth is larger than at the upper turning depth due to the larger horizontal extent of the ray loops below rather than above the channel axis. As for the comparison with the wave-theoretic VTSKs, a similar behavior with increasing range is observed as in the case of the linear profile. At the shortest propagation range of 50 km, there is a large deviation around the lower turning depth, with the wave-theoretic TSK extending by more than 700 m below the ray-theoretic turning depth. As the range increases, the differences become smaller and for ranges above 500 km the two kernels are very close to each other.

IV. DISCUSSION AND CONCLUSIONS

The focus of this work is to compare TSKs and VTSKs in ocean acoustic propagation calculated using different approximations and to study the effects of increasing range. Wave-theoretic sensitivity kernels calculated in two and three dimensions using normal modes and in two dimensions using the parabolic approximation were compared against each other and also against ray-theoretic sensitivity kernels.

The 2D and the 3D TSKs concentrate about the corresponding eigenrays but they are different. For short and medium propagation ranges, the 3D TSK exhibits a near-zero-sensitivity core in the vicinity of the eigenray, with maximum extent near the turning points, surrounded by a negative-sensitivity domain, followed by positive sensitivity, etc. The 2D TSK, on the other hand, exhibits uniform negative sensitivity near the eigenray; i.e., it lacks the zero-sensitivity core. Further, the widths of the two TSKs in the vertical source-receiver plane are different. These differences appear to average out when the TSKs are integrated in cross-range or in the horizontal, i.e., in the case of 2D and 3D HTSKs and VTSKs. The zero-sensitivity core cannot be identified at all in the 3D VTSK or HTSK. In fact, this core in the 3D TSK comes into effect only for sound-speed perturbations of small horizontal scale of the order $O(100\ \mathrm{m})$, comparable to the cross-range extent of the core. In reality, much larger horizontal scales characterize the sound-speed perturbations in the ocean; taking integrals over such scales results in disappearance of the zero-sensitivity effect. In this connection, the wave-theoretic treatment of travel-time observables using a 2D or 3D approach is not expected to be significantly different.

As the range increases, the 3D TSK initially contracts and then it slowly expands in the vertical; this behavior is best observed at midrange between source and receiver where the TSK extent is largest. On the other hand, while for short ranges the wave-theoretic TSKs and VTSKs extend well beyond the corresponding ray-theoretic lower turning depth at the low frequency considered (100 Hz), for long ranges they hardly surpass the ray-theoretic depth limits. A similar behavior characterizes the vertical extent of the corresponding Fresnel-volumes.

As for the horizontal cross-range direction, the range increase causes the 3D TSK to expand monotonically at the turning depth, similar to the behavior of the free-space Fresnel zone. This is anticipated for the range-independent environment under study, since there is no horizontal refraction. Concerning the zero-sensitivity core of the 3D TSK, the results presented here for the lower turning points show that for short propagation ranges the core has circular symmetry about the eigenray. As the range increases, the core deviates from the circular symmetry by contracting in the vertical and expanding in the horizontal cross-range direction, and at long ranges it disappears. At the turning points, the horizontal cross-range extent of the negative-sensitivity core of the wave-theoretic TSKs prevails over the corresponding vertical extent.

The deviation between wave-theoretic and ray-theoretic VTSKs is largest for the short/medium ranges. While ray-theoretic VTSKs are strictly negative between the upper and lower turning depths, the low-frequency short/medium-range wave-theoretic VTSKs extend significantly beyond the turning depths, especially beyond the lower turning depth, and they even become positive over particular depth intervals. As the propagation range increases, these differences become smaller, and finally long-range wave-theoretic and ray-theoretic VTSKs are nearly identical. One would typically expect such a behavior from a frequency increase, since ray theory is a high-frequency asymptotic approximation. However, the frequency considered here is low (100 Hz), and despite this fact the range increase brings the wave-theoretic VTSK close to the ray-theoretic one.

To check the above wave-theoretic behavior, additional VTSK calculations were carried out based on the parabolic approximation. The VTSKs obtained from the two alternative wave-theoretic approaches (normal-mode and parabolic approximation) compare favorably to each other for longer ranges. This further supports the conclusion that for long-range propagation ray-theoretic VTSKs can be used even at low frequencies. Further, a significant advantage of the parabolic approximation is that it allows for TSK calculation in range-dependent environments.

The results presented here offer support to the use of ray theory for the interpretation of low-frequency long-range transmissions. The convergence of the negative core of the TSK at the turning points toward the eigenray (in the vertical), combined with its expansion in the horizontal similar to the free-space Fresnel zone, as the range increases, indicates that wave-theoretic travel times are sensitive to sound-speed changes on the eigenrays and their horizontal (cross-range) neighborhood over Fresnel scales. This is further reinforced by the convergence with increasing range of wave-theoretic VTSKs toward the corresponding ray-theoretic kernels.

Topics for future work include the study of TSK and VTSK behavior in dependence from the stability parameter α,[20] as well as the physical explanation of the convergence of the wave-theoretic VTSKs to the ray-theoretic ones with increasing range. The agreement between wave- and ray-theoretic VTSKs at the upper turning point, with simultaneous disagreement at the lower turning point, in the case of the temperate sound-speed profile for short ranges (Fig. 12) is possibly a good starting point for seeking a physical interpretation.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for helpful comments and suggestions. This work was supported by ONR, in the framework of Contract Nos. N00014-06-1-0407, N00014-08-0840, N00014-03-1-0182, and N00014-07-1-0739.

[1] H. Marquering, G. Nolet, and F. A. Dahlen, "Three-dimensional waveform sensitivity kernels," Geophys. J. Int. **132**, 521–534 (1998).
[2] H. Marquering, F. A. Dahlen, and G. Nolet, "Three-dimensional sensitivity kernels for finite-frequency traveltimes: The banana-doughnut paradox," Geophys. J. Int. **137**, 805–815 (1999).
[3] F. A. Dahlen, S.-H. Hung, and G. Nolet, "Fréchet kernels for finite-frequency traveltimes—I. Theory," Geophys. J. Int. **141**, 157–174 (2000).
[4] S.-H. Hung, F. A. Dahlen, and G. Nolet, "Fréchet kernels for finite-frequency traveltimes—II. Examples," Geophys. J. Int. **141**, 175–203 (2000).
[5] J. Tong, F. A. Dahlen, G. Nolet, and H. Marquering, "Diffraction effects upon finite-frequency travel times: A simple 2-d example," Geophys. Res. Lett. **25**, 1983–1986 (1998).
[6] M. Born, "Quantum mechanics of impact processes," Z. Phys. **38**, 803–827 (1926).
[7] J. R. Taylor, *Scattering Theory* (Wiley, New York, 1972).
[8] G. A. Athanassoulis and E. K. Skarsoulis, "Arrival-time perturbations of broadband tomographic signals due to sound-speed disturbances. A wave-theoretic approach," J. Acoust. Soc. Am. **97**, 3575–3588 (1995).
[9] E. K. Skarsoulis, G. A. Athanassoulis, and U. Send, "Ocean acoustic tomography based on peak arrivals," J. Acoust. Soc. Am. **100**, 797–813 (1996).
[10] E. K. Skarsoulis and B. D. Cornuelle, "Travel-time sensitivity kernels in ocean acoustic tomography," J. Acoust. Soc. Am. **116**, 227–238 (2004).
[11] W. H. Munk, P. F. Worcester, and C. Wunsch, *Ocean Acoustic Tomography* (Cambridge University Press, New York, 1995).
[12] W. H. Munk and C. Wunsch, "Ocean acoustic tomography: A scheme for large scale monitoring," Deep-Sea Res., Part A **26A**, 123–161 (1979).
[13] C. A. Boyles, *Acoustic Waveguides. Applications to Oceanic Science* (Wiley, New York, 1984).
[14] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, Woodbury, NY, 1994).
[15] M. D. Collins, "A split-step Pade solution for the parabolic equation method," J. Acoust. Soc. Am. **93**, 1736–1742 (1993).
[16] M. D. Collins, "Generalization of the split-step Pade solution," J. Acoust. Soc. Am. **96**, 382–385 (1994).
[17] M. D. Collins, "A self-starter for the parabolic equation method," J. Acoust. Soc. Am. **92**, 2069–2074 (1992).
[18] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-relativistic Theory* (Pergamon, Oxford, 1977).
[19] Yu. A. Kravtsov and Yu. I. Orlov, *Geometrical Optics of Inhomogeneous Media* (Springer, Berlin, 1990).
[20] I. I. Rypina and M. G. Brown, "On the width of a ray," J. Acoust. Soc. Am. **122**, 1440–1448 (2007).
[21] B. D. Cornuelle and P. F. Worcester, "Ocean acoustic tomography: Integral data and ocean models," in *Modern Approaches to Data Assimilation and Ocean Modelling*, Elsevier Oceanography Series, edited by P. Malanotte-Rizzoli (Elsevier, New York, 1996), pp. 97–115.

Effects of sea-surface conditions on passive fathometry and bottom characterization

Steven L. Means[a)]
Naval Research Laboratory, Code 7120, 4555 Overlook Avenue SW, Washington, DC 20375

Martin Siderius
Maseeh College of Engineering and Computer Science, Portland State University, P.O. Box 751, Portland, Oregon 97207

(Received 6 March 2009; revised 6 August 2009; accepted 6 August 2009)

Recently, a method has been developed that exploits the correlation properties of the ocean's ambient noise to measure water depth (a passive fathometer) and seabed layering [M. Siderius *et al.*, J. Acoust. Soc. Am. **120**, 1315–1323 (2006)]. This processing is based on the cross-correlation between the surface noise and the echo return from the seabed. To quantitatively study the dependency between processing and environmental factors such as wind speed, measurements were made using a fixed hydrophone array while simultaneously characterizing the environment. The measurements were made in 2006 in the shallow waters (25 m) approximately 75 km off the coast of Savannah, GA. A Navy tower about 100 m from the array was used to measure wind speed and to observe the sea-surface using a video camera. Data were collected in various environmental conditions with wind speeds ranging from 5 to 21 m/s and wave heights of 1–3.4 m. The data are analyzed to quantify the dependency of passive fathometer results on wind speeds, wave conditions, and averaging times. One result shows that the seabed reflection is detectable even in the lowest wind conditions. Further, a technique is developed to remove the environmental dependency so that the returns estimate seabed impedance. [DOI: 10.1121/1.3216915]

PACS number(s): 43.30.Pc, 43.30.Wi, 43.60.Pt [AIT] Pages: 2234–2241

I. INTRODUCTION

In most sonar signal processing applications, ambient noise is considered a negative entity. Generally, ideal conditions for sonars are those where the ambient noise is very low. However, in recent years techniques have been developed to exploit the ambient noise field for useful applications.[1,2] Recently a new method of processing ambient noise measurements has allowed for the extraction of information about the sea bottom.[3] Specifically, this new method makes it possible to measure water depth (a fathometer) and seabed layering using just the ambient noise field. There are several good reasons to study techniques that use ocean noise rather than sound projectors as with traditional active sonar methods. For one, the controversy over the effects of man-made sounds on marine life highlights the need for environmentally friendly remote sensing tools such as these ambient noise systems. Further, using ambient noise rather than high-powered, man-made sound sources simplifies the measurements.

The passive fathometer methodology developed by Siderius *et al.*[3] exploits processing the coherent components of the noise field. The passive fathometer is based on the cross-correlation between the surface "signal," generated by breaking waves, and the echo return from the seabed. The "signal level" depends on the nature of the breaking waves, which in turn depends on other environmental factors such as wind speed and fetch. For practical applications with, for example, autonomous systems, it is critical to understand the parameters important to the signal processing, for example, averaging times, time snapshot size, and required sea-state. To study these parameters quantitatively, a fixed hydrophone array together with careful measurements of the environment is essential. A moving system has too many variables changing (such as water depth or bottom type) to isolate the effects of the surface conditions and the signal processing so that their dependencies can be studied.

A number of questions are addressed through the analysis of the passive fathometer response with simultaneous wind speed measurements and video of the sea-surface conditions at a fixed array. For example, what are the minimum wind speed (or sea-state) conditions required and what is the dependency of the response on wind or sea-state conditions? The data considered in this article were taken from a long-term deployment that allowed a wide variety of conditions to be studied. A second goal of this work is to describe how the passive fathometer return can provide a quantitative measure of the impedance contrast between the water and the seabed layers. This provides a very simple yet useful measurement for identifying the seabed type (e.g., gravel, sand, mud, etc.). To accomplish this the processing needs to be self-calibrating to remove any dependency on wind speed or sea-state. A deployed system would not be nearly as useful if the impedance estimate required an ancillary wind speed or sea-state measurement.

This paper is organized as follows: The experiment's location, equipment used, and the measured environmental

[a)]Author to whom correspondence should be addressed. Electronic mail: steve.means@nrl.navy.mil

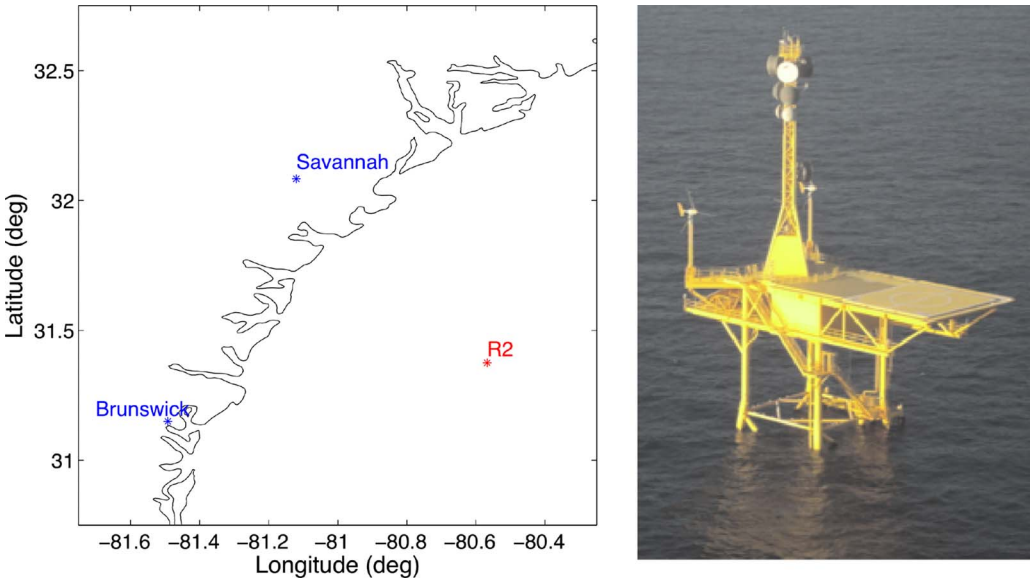


FIG. 1. (Color online) Location of TACTS off-shore range and image of R2 tower.

parameters are presented in Sec. II. In Sec. III, an overview of the processing used to obtain the passive fathometer's time-series response is given. The effects of environmental conditions on the fathometer's uncalibrated response are presented and discussed in Sec. IV. An analysis of optimal processing parameters for improved detection of bottom features is presented in Sec. V. A newly developed algorithm to calibrate the fathometer response so that the magnitude of the response from a given bottom feature represents the reflection loss is developed and investigated within Sec. VI. A summary of the research findings concludes this paper in Sec. VII.

II. EXPERIMENTAL MEASUREMENTS

In January 2006 a 32-hydrophone, three-nested aperture array was deployed near an offshore platform (see Fig. 1). The platform, R2, is one of a range of offshore towers operated by the Navy as a part of a Tactical Air Combat Training System. The tower is located in the shallow waters (25 m) approximately 75 km off the coast of Savannah, GA and extends 50 m above the water surface. The tower is equipped to supply power through solar panels, wind turbines, and a diesel generator. Additionally, it is equipped with two-way microwave communication back to shore, which allowed for long-term measurements while controlling the data acquisition from land via the internet.

The array had hydrophone spacings of 1, 0.5, and 0.25 m yielding design frequencies of 750, 1500, and 3000 Hz, respectively. Figure 2 illustrates the 15-dB down end-fire beam radius of the innermost aperture on the ocean bottom, used in the analysis presented here.

A high-resolution video camera was installed at the top of the tower (~50 m above water surface). It allowed time-synchronized monitoring of the ocean's surface above the vertical acoustic array. The camera and lens were calibrated using the camera calibration toolbox for MATLAB (Ref. 4) software so that the obtained images could then be georectified to obtain an overhead view of the surface above the array.

In addition to the acoustic and video measurements, the Skidaway Institute of Oceanography, as a component of the South Atlantic Bight Synoptic Offshore Observational Network, maintains a suite of environmental sensors on the tower. These provide both meteorological and oceanographic measurements, which are available from Skidaway's website.[5] Of interest here are measurements of wind speed and wave height. Measurements of the tide were also available; however, due to the use of a bottom-fixed array, it had no effect on the results reported here.

III. PROCESSING

The passive fathometer is based on the cross-correlation of the surface noise generated by breaking waves and the echo return from the seabed. For a good portion of the frequency band, except lower frequencies dominated by shipping (~20–200 Hz), the breaking waves are commonly the predominant source of ambient noise (up to ~30 kHz). The passive fathometer processing was developed by Siderius *et al.*[3] and since the original introduction, a number of efforts[6,7] have extended and refined the methodology and improved

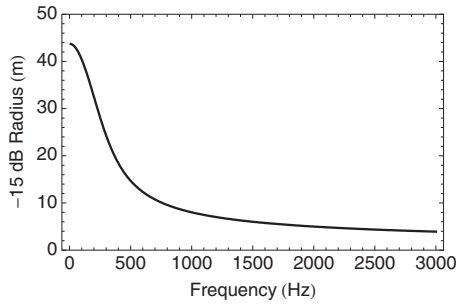


FIG. 2. End-fire beam radius (−15 dB) of the innermost array aperture at the ocean bottom. This gives an idea of the bottom surface patch size interrogated as a function of frequency.

the understanding. This work built on the seminal work of Rickett and Claerbout[8] and Weaver and Lobkis[9,10] in seismic and ultrasonics as well as Kuperman and co-workers for underwater acoustics.[11–14]

The simplest formulation starts with correlation between a beam in the direction toward the surface with the beam toward the seabed. These beams are formed using an array such as that described in Sec. II. The hydrophone data at frequency ω are written as a column vector $\mathbf{d}=[d_1,d_2,\ldots,d_M]$ for the M hydrophones. In conventional beamforming, the weight for the mth hydrophone steered at 90° (in direction toward the surface) is written as

$$w_m = e^{i(ma\omega/c)}, \tag{1}$$

where a is the distance between the equally spaced hydrophones and c is the sound speed in the water (around 1500 m/s). If the surface steering weights are written as a column vector, $\mathbf{w}=[w_1,w_2,\ldots,w_M]$, the beam directly toward the surface, B_{up}, can be written as

$$B_{\mathrm{up}} = \mathbf{w}^{\dagger}\mathbf{d}, \tag{2}$$

where † represents the conjugate transpose operation. The steering weights toward the seabed (at −90°) are just the conjugate of the weights steered toward the surface. The beam steered directly toward the seabed is then

$$B_{\mathrm{dn}} = \mathbf{w}^{\mathbf{T}}\mathbf{d}, \tag{3}$$

where $\mathbf{T}$ represents the transpose operation without conjugation. The correlation of the surface steered beam with the seabed steered beam is

$$C = B_{\mathrm{up}}B_{\mathrm{dn}}^{*} = (\mathbf{w}^{\dagger}\mathbf{d})(\mathbf{w}^{\mathbf{T}}\mathbf{d})^{*} = \mathbf{w}^{\dagger}\mathbf{d}\mathbf{d}^{\dagger}\mathbf{w}^{*} = \mathbf{w}^{\dagger}\mathbf{K}\mathbf{w}^{*}, \tag{4}$$

where the cross-spectral density matrix (CSDM), $\mathbf{K}$, is identified as a time average of $\mathbf{d}\mathbf{d}^{\dagger}$ and * indicates a conjugation. Note that if $\mathbf{w}^*$, in Eq. (4), is replaced with $\mathbf{w}$ one obtains the expression for a beam steered toward the surface as opposed to a cross-correlation between upward and downward beams. With the given expression, the CSDM can be formed over as many snapshots of data, $\mathbf{d}$, as needed to obtain the desired averaging. The number of snapshots needed is one of the topics of Sec. V.

An improved fathometer response can be achieved by using adaptive beamforming, or specifically, minimum variance distortionless response (MVDR).[15] MVDR is useful to suppress the energy coming from directions other than that of interest. In this case there is significant energy coming near horizontal that is of no interest (i.e., snapping shrimp colony on the R2 tower) for the passive fathometer processing. To adaptively beamform, the MVDR steering weights, $\mathbf{w_A}$, are computed, according to Burdic,[16] as

$$\mathbf{w_A} = \frac{\mathbf{K}^{-1}\mathbf{w}}{\mathbf{w}^{\dagger}\mathbf{K}^{-1}\mathbf{w}}. \tag{5}$$

The MVDR correlation at frequency ω is

$$C = \mathbf{w_A}^{\dagger}\mathbf{K}\mathbf{w_A}^{*}. \tag{6}$$

The time-series passive fathometer response is the inverse Fourier transform of C or $r(t)=\mathcal{F}^{-1}\{C(\omega)\}$. Strictly, this expression is fine if detecting the seabed and layering are all that is of interest. However, if one wants the impulse response of the seabed, $\grave{r}(t)$, then differentiation with time is needed as described by Harrison and Siderius[6] and Roux *et al.*[12] The Fourier transform of the impulse response, $\grave{r}(t)$, is the reflection coefficient, $R(\omega)$.[17] An estimate for the impulse response, $\tilde{r}(t)$, is then

$$\tilde{r}(t) = N\frac{d}{dt}r(t), \tag{7}$$

where N is an unknown normalization constant, as derived by Harrison and Siderius,[6] that involves several terms, including the beam width, integration time, and the standard deviation of the noise (related to the sea-state). Some of these terms, such as those that depend on the exact surface conditions, make estimating this factor difficult.

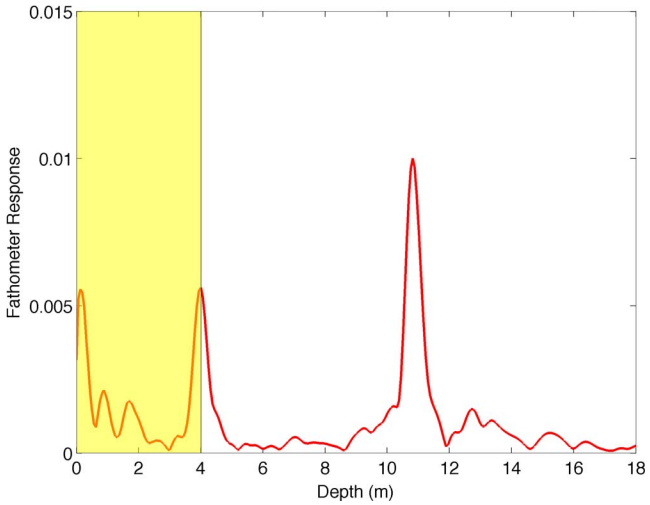


FIG. 3. (Color online) Uncalibrated fathometer response referenced to the top hydrophone at 10.96 m above ocean bottom as measured at deployment. The shaded area represents the two-way travel time over the length of the array and the response within it may be considered as a processing artifact.

Figure 3 shows a typical, uncalibrated, fathometer response, $r(t)$, for the experiment environment that is referenced to the topmost hydrophone of the innermost aperture. The time axis has been converted to distance using the two-way travel time assuming a sound speed of 1500 m/s. The initial response within the first 4 m (shaded box) is a processing artifact that corresponds to the length of the array.[3,7] The peak at ~11 m is the response due to the bottom and is in good agreement with the known bottom depth (from the topmost hydrophone) of 10.96 m.

IV. EFFECTS OF SEA-SURFACE CONDITIONS

Presumably any passive fathometry systems developed in the future will be required to operate in a variety of environmental conditions; thus it is of interest to analyze the impact of the conditions on the processing results. Although of little interest for actual fathometry, the fixed location of the array used here is ideal for such a study. Data sets from 2 days were selected to investigate the environmental effects on the fathometry processing. The first set was acquired on January 14, 2006 over roughly an 8-h time period. The en-

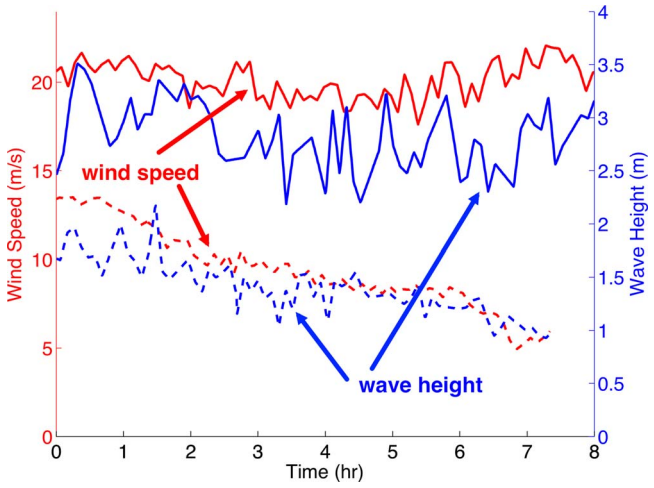


FIG. 4. (Color online) Wind speeds and wave heights for the 2 days analyzed within the study. Solid lines represent wind speeds and wave heights acquired on Julian day 14. Dashed lines were obtained on Julian day 81, in which wind speeds and wave heights declined throughout the day.

vironmental conditions were relatively constant with high winds (~20 m/s) and wave heights of ~3 m (see Fig. 4). The second data set was acquired during a time period (March 22, 2006) in which the wind speeds and wave heights dropped over the duration of several hours.

Initial development of this methodology[3] plausibly assumed that breaking waves were the source that made the processing feasible. This assumption is proven correct here via video images of the sea surface recorded simultaneously with the acoustic data. Figures 5 and 6 show the time-synchronized video images (georectified), upward end-fire beam spectrograms, and normalized fathometer responses (see Sec. V for normalization process) in the absence and in the presence of a breaking wave, respectively. The outlined windows in the spectrogram figures represent the 10-s averaging window, and the video snapshots correspond to its leading edge. It is seen in Fig. 5 that in the absence of a

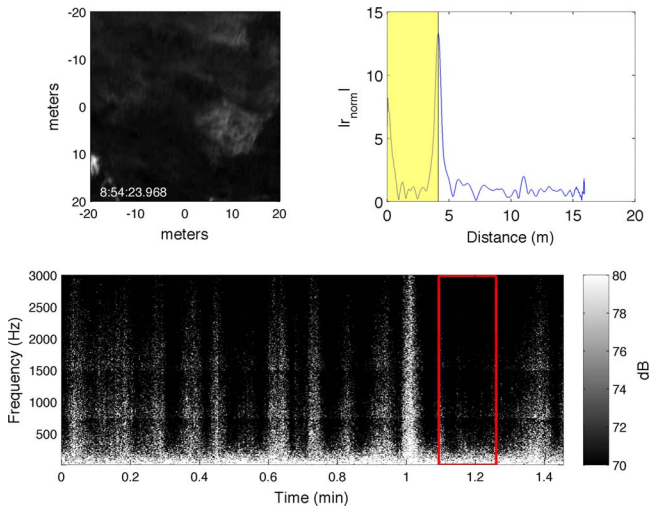


FIG. 5. (Color online) Time-synchronized images of ocean surface video, end-fire beam acoustic array reception, and normalized fathometer response. In the absence of breaking waves (video and acoustic) within a 10-s averaging time window, no fathometer response is observed at the known bottom depth (10.96 m).

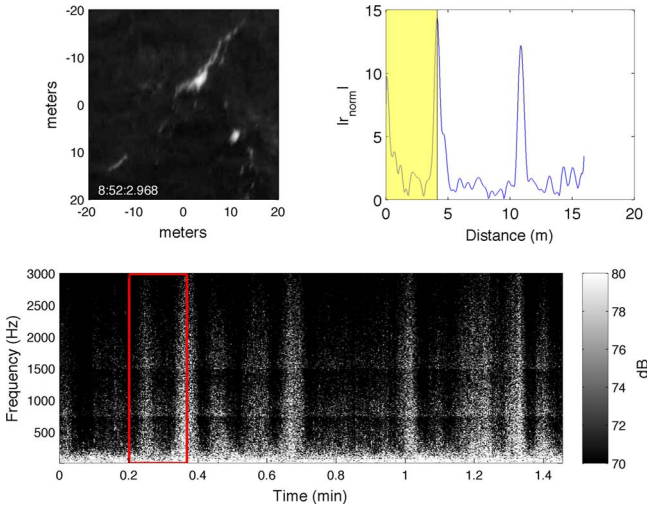


FIG. 6. (Color online) Time-synchronized images of video, end-fire acoustic array reception, and normalized fathometer response. In the presence of breaking waves (video and acoustic) within a 10-s averaging time window, a peak in the fathometer response is observed at the known bottom depth (10.96 m).

breaking wave, within the end-fire beam pattern and the processing averaging time window (10 s), no fathometer response is seen at the known bottom depth (~11 m). However, when a breaking wave does occur overhead of the array (see Fig. 6), a strong peak is seen in the fathometer response at the known bottom depth.

In an effort to investigate the effect of wind speed and wave height on the fathometer response, the amplitude of the response at the known bottom location was examined. Figures 7 and 8 show log-log plots of the peak of the unnormalized fathometer response (using an 80-s averaging time) as functions of wind speed and wave height, respectively. The data points plotted as triangles correspond to Julian day 14, in which wind speed remained constant, and the asterisks

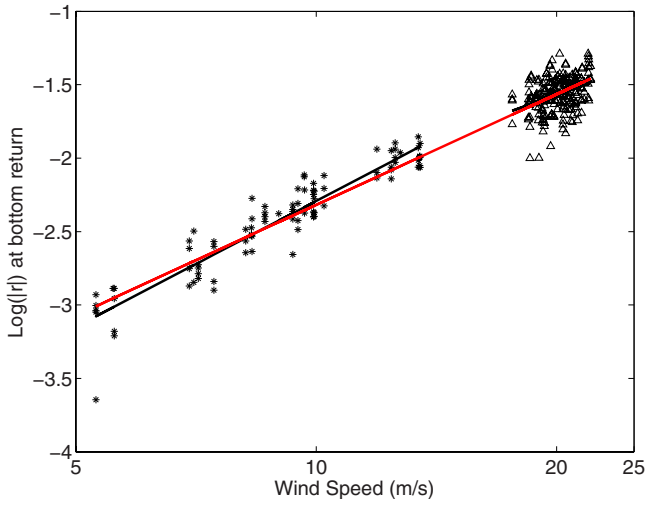


FIG. 7. (Color online) Relationship between wind speeds and the magnitude of raw fathometer response at the bottom return (80-s averaging time). The asterisks represent data taken on Julian day 81, and the linear best fit has a slope of 2.842 and wind speed correlation of 0.9395. The triangles represent data taken on Julian day 14, and the linear best fit has a slope of 1.9342 and a correlation of 0.3584. The linear best fit for the 2 days of data has a slope of 2.503 and a correlation of 0.9445.

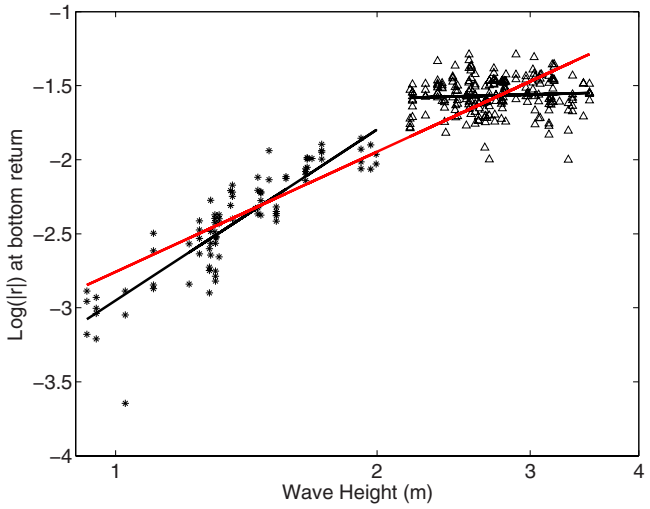


FIG. 8. (Color online) Relationship between wave heights and the magnitude of raw fathometer response at the bottom return (80-s averaging time). The asterisks represent data taken on Julian day 81, and the linear best fit has a slope of 2.8482 and wind speed correlation of 0.8808. The triangles represent data taken on Julian day 14, and the linear best fit has a slope of 0.1519 and a correlation of 0.0670. The linear best fit for the 2 days of data has a slope of 2.6959 and a correlation of 0.9081.

plot data taken during the declining winds of Julian day 81. A best-fit line has been inserted for both data sets. In comparing the two figures and linear best fits, it is seen that the fathometer response is better correlated with wind speed than wave height. [Correlations are 0.9445 (JD 14 and 81) with wind speed versus 0.9081 (JD 14 and 81) with wave height.] The linear nature of the relationship between the fathometer response with the wind speed, as plotted, seems to follow observed relationships between noise level and wind speed.[18] Mixed-sea conditions were observed (via video) during much of the acquisition on Julian day 14, which is a plausible explanation for the higher variability in the fathometer response as a function of both wind speed and wave height.

V. NORMALIZATION, OPTIMIZATION, AND DETECTABILITY

In practice there are a few signal processing parameters which can be adjusted to optimize the fathometer processing for a given environment. Two of importance are the length of fast Fourier transform (FFT) (snapshot size) and the averaging time (number of snapshots). First of all, the FFT length must be selected so that sufficient travel time is allowed for the propagation to and from the bottom, and any sub-bottom features of interest.

In addition to adjusting processing parameters, one may also choose to normalize the fathometer response. This allows comparing one result to another or determining optimal performance for given conditions. A better understanding of the detectability of bottom and, presumably, sub-bottom returns may be gained by normalizing the fathometer response. The normalization chosen here is the mean of the "noise" background between the initial processing artifacts (i.e., travel along the length of the array) and the response due to the bottom return. This occurs over depths between ~5 and 10 m, as seen in Fig. 3.

Figures 9(a) and 9(b) show the magnitude of the peak in the normalized fathometer response, $|r_{\mathrm{norm}}|$, as a function of averaging time for different wind speeds with FFT lengths of 2.73 and 0.17 s, respectively. Each curve represents the mean of 15–20 time segment samples (with averaging times as indicated) within an hour time period. It is evident that for longer FFT lengths, Fig. 9(a), a longer averaging time is necessary to obtain a more distinct peak. However, for the shorter FFT lengths, Fig. 9(b), the peak's magnitude begins to plateau at shorter averaging times. Additionally, lower wind speeds (though, still with wave breaking) and longer FFT lengths may require long averaging times to observe the bottom return.

Figure 10 shows the uncalibrated fathometer response amplitude at the bottom return and the normalization factor

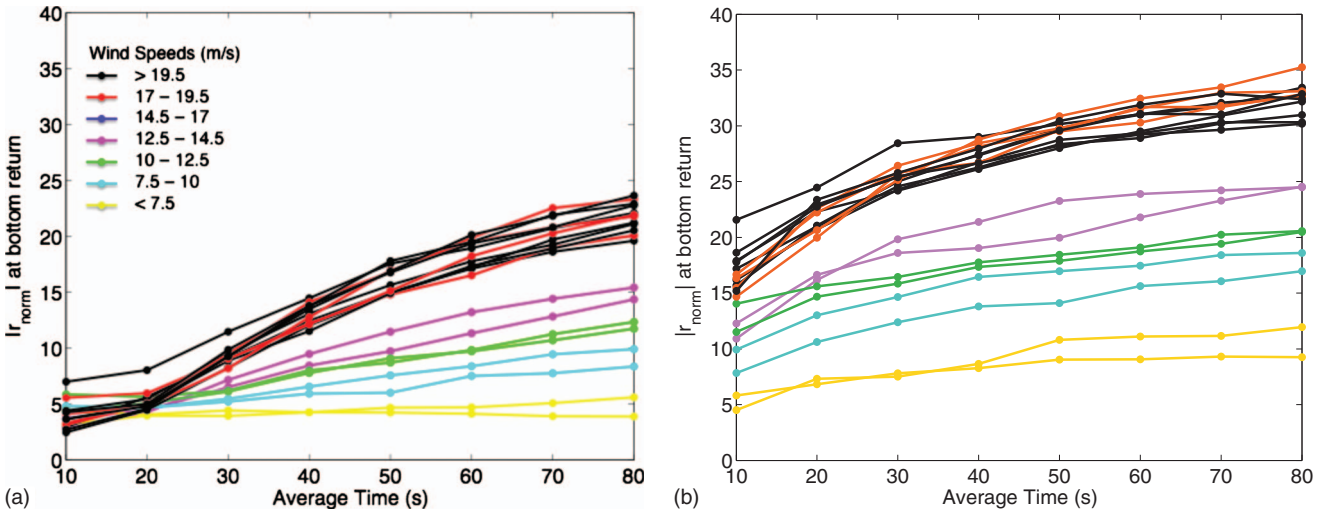


FIG. 9. Magnitude of normalized fathometer response at bottom return as a function of averaging time using adaptive beamforming for a range of wind speeds. Panel (a) was processed with an FFT length of 2.73 s; panel (b) was processed with an FFT length of 0.17 s. It is evident that shorter FFT lengths achieve higher detectability with shorter averaging times, which would provide better resolution in a drifting fathometer system in ocean conditions with less frequent wave breaking.

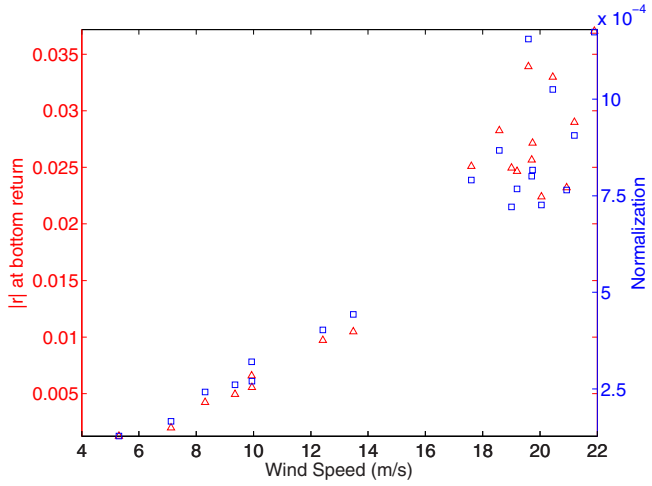


FIG. 10. (Color online) Magnitude of uncalibrated fathometer response peaks (triangles) and normalization factors (squares) as a function of wind speed using adaptive beamforming and 80-s averaging time with FFT length of 0.17 s.

as a function of wind speed. It is seen that as wind speed increases the peaks become more detectable as the ratio of the raw response and the normalization factor becomes larger.

Although it is plausible that the normalization used here would remove the dependence on wind speed (or wave height) on the fathometer response, it proves not to be the case. Figure 11 shows the magnitude of the normalized fathometer response at the bottom return along with the wind speeds for the two time periods. It is seen that the response still tracts with the wind speed.

As discussed in Sec. III, adaptive beamforming is used here primarily. However, to illustrate the improvement this allows, Fig. 12 shows the magnitude of the fathometer response as a function of averaging time and wind speeds when conventional beamforming is employed, rather than

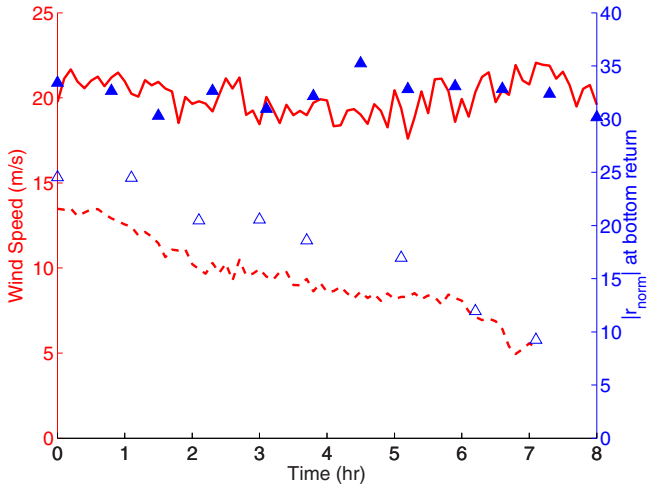


FIG. 11. (Color online) Wind speed and magnitude of normalized fathometer response as a function of time with FFT length of 0.17 s and an averaging time of 80 s. Solid line and triangles represent wind speeds and power reflection losses acquired on Julian day 14. Dashed lines and open triangles were obtained on Julian day 81, in which wind speeds and wave heights declined throughout the day. The normalized fathometer response tracks with the wind speed.

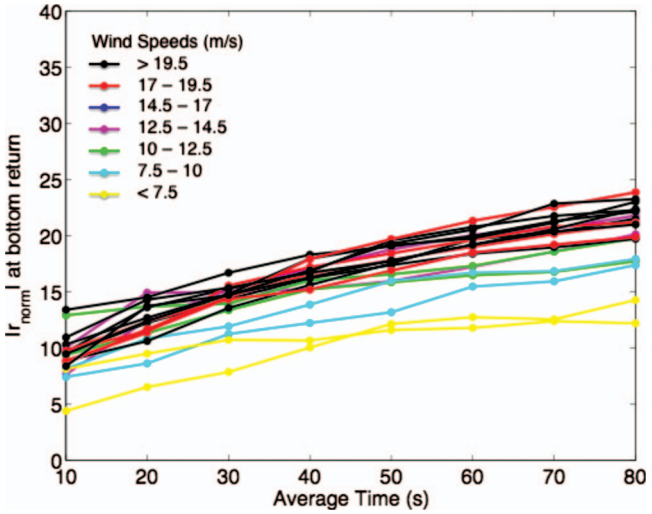


FIG. 12. Magnitude of normalized fathometer response as a function of averaging time with FFT length of 0.17 s using conventional beamforming.

the adaptive beamforming which is used in Fig. 9(b). It is seen that for conventional beamforming, additional averaging time would be necessary to maximize the detectability of returns within the fathometer response.

VI. CALIBRATED RESPONSE

One may also choose to normalize the fathometer response so that the return is a direct estimate of the bottom impulse response $\tilde{r}(t) \approx \dot{r}(t)$. This will be referred to as the calibrated passive fathometer response and ideally is independent of wind speed or sea-state. With $r(t)$, the value at any peak represents the reflection coefficient at that interface. Recall from Eq. (7) that the estimate of the impulse response $\tilde{r}(t)$ is related to the cross-correlation through a time-derivative and an unknown factor N. To determine N, the relationship developed by Harrison and Simons[2] for estimating the magnitude-squared reflection coefficient in terms of the beam powers can be used,

$$|\tilde{R}|^2 = \frac{|B_{\mathrm{up}}|^2}{|B_{\mathrm{dn}}|^2}. \tag{8}$$

The power in the time and frequency domains is the same; therefore,

$$\int_{-\infty}^{\infty} |\dot{r}(t)|^2 dt = \int_{-\infty}^{\infty} |R(\omega)|^2 d\omega. \tag{9}$$

The estimated value for N is determined using Eqs. (8) and (7),

$$N = \sqrt{\frac{\int_{-\infty}^{\infty} |\tilde{R}(\omega)|^2 d\omega}{\int_{-\infty}^{\infty} |(d/dt) r(t)|^2 dt}}. \tag{10}$$

In practice, the integration limits used are based on the available bandwidth and array geometry (i.e., hydrophone spacing). The reflection coefficient between two media at normal incidence is defined as

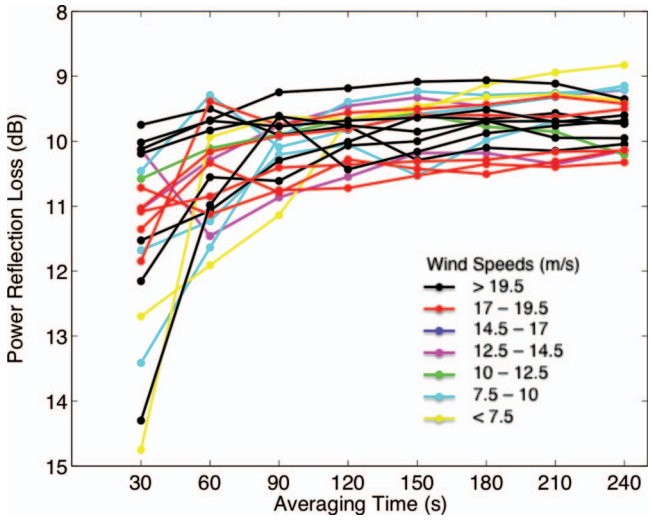


FIG. 13. Calibrated fathometer response (power reflection loss), via coherent processing, as a function of averaging time. This is in good agreement with the mean power reflection loss of 7.8 dB at vertical incidence, as determined by Eq. (8), for frequencies between 500 and 2800 Hz.

$$R = \frac{Z_2 - Z_1}{Z_2 + Z_1}, \tag{11}$$

where Z_1 and Z_2 are the impedance values in the two media (e.g., water and seabed) which are $Z_1 = c_1\rho_1$ and $Z_2 = c_2\rho_2$ for sound speeds, c_1 and c_2, and densities, ρ_1 and ρ_2. The time domain version of the reflection coefficient, $\tilde{r}(t)$, is useful since the peak values give the impedance at the water-seabed interface as well as between sub-bottom layers. Using Eq. (10) in Eq. (7) results in a type of calibration for the passive fathometer time-series that should not depend on factors such as integration time and sea-state. Figure 13 shows the calibrated fathometer response peak (magnitude-squared in decibels) as a function of averaging time and wind speeds. This represents the reflection loss, and the figure shows reasonable stability in the estimate over a large range of wind speeds once the averaging is above about 1 min.

This case is somewhat trivial since there is only one peak in the fathometer return; however, in principle this should provide the impedance contrast value for additional layers using this process. The bottom loss (i.e., $-10\log(|R|^2)$) at vertical incidence was also calculated using the frequency domain calculation given by Eq. (8). This was integrated over frequencies from 500 and 2800 Hz to produce a loss estimate of 7.8 dB. If there were significant layering, however, there would be a complicated interference pattern in the frequency domain reflection coefficient. Integration of the reflection coefficient over frequency would not represent the loss at any of the individual layers, just the integrated loss over frequency through the entire seabed.

For this impedance estimating methodology to be useful in practical applications, it must yield values which are invariant to environmental conditions. Figure 14 shows the magnitude of the calibrated fathometer response (in decibels) for different times and wind speeds. It is seen that the reflection loss varies by less than 0.3 dB when an averaging time of 80 s is used.

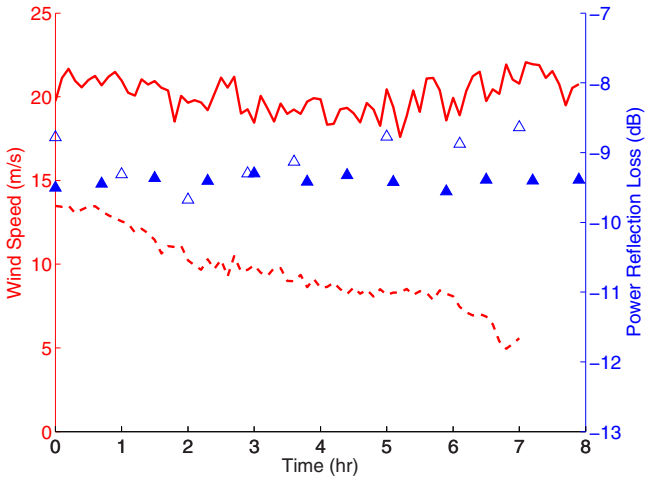


FIG. 14. (Color online) Wind speed and magnitude of calibrated fathometer response as a function of time with FFT length of 0.17 s and an averaging time of 240 s. The solid line is wind speed data acquired on Julian day 14 and solid triangles are the corresponding reflection loss value. Dashed line is wind speed for Julian day 81, with open triangles indicating the corresponding reflection loss values.

Comparison of the fathometer response can be made with sediment measurements made during the Tactical Air Combat Training System (TACTS) tower construction (see Table I).[19] The estimated sound speeds and densities are obtained from an APL-UW handbook[20] using the description of the layers within the core sample. Previous analysis by Siderius *et al.*[3] of data obtained above a softer bottom has shown that the passive fathometer was capable of detecting sub-bottom layering within the sediment to 25 m beneath the bottom. Although sub-bottom layering is present in the core samples drawn prior to tower construction, no sub-bottom layering is observed via the passive fathometer. It is likely that the large impedance mismatch, due to the sandy bottom, prevents the detection of sub-bottom layering. However, note that the impedance from the table values [Layer 1 reflection loss: 8.6–12.6 dB using Eq. (11)] is in relatively good agreement with the reflection loss values estimated here from the calibrated passive fathometer response.

VII. SUMMARY AND CONCLUSIONS

The main emphasis of this paper has been to determine the effect of environmental conditions on passive fathometer

TABLE I. Core sample description and estimated sound speeds and densities.

Layer	Description	Depth (m)	Est. sound speed[a] (m/s)	Est. density (kg/m^3)
1	Gray calcareous fine to medium sand	0.0	1660–1767	1451–1845
2	Greenish gray carbonate silty fine sand	4.88	1660	1451
3	Sandy clay	14.0	1477	1147
4	Hard silty clay	24.70	1473	1146
5	Hard calcareous olive gray sandy clay	76.22	1477	1147

[a]Reference 20.

processing techniques. The analysis was focused on determining optimal processing parameters over a range of wind speeds and sea-states to aid the development of practical passive fathometer systems. It was determined that for a given depth of interest, shorter FFT lengths yield more detectable bottom returns with less averaging time. Thus, for practical moving systems, higher-resolution fathometry can be performed when shorter FFT lengths are used. It was also observed that the adaptive beamforming methods (e.g., MVDR) yielded increased performance. Additionally, a new self-calibrating methodology was proposed such that the magnitude of fathometer response peaks yield estimates of reflection coefficients of bottom and sub-bottom interfaces. It was shown that the proposed technique yielded valid results independent of wind speed.

ACKNOWLEDGMENTS

This work was supported by Office of Naval Research base funding at the Naval Research Laboratory.

[1]M. J. Buckingham, B. V. Berkhout, and S. A. L. Glegg, "Imaging the ocean with ambient noise," Nature (London) **356**, 327–329 (1992).
[2]C. H. Harrison and D. G. Simons, "Geoacoustic inversion of ambient noise: A simple method," J. Acoust. Soc. Am. **112**, 1377–1389 (2002).
[3]M. Siderius, C. H. Harrison, and M. B. Porter, "A passive fathometer technique for imaging seabed layering using ambient noise," J. Acoust. Soc. Am. **120**, 1315–1323 (2006).
[4]http://www.vision.caltech.edu/bouguetj/calib_doc (Last viewed 9/11/2009).
[5]http://www.skio.usg.edu/Skioresearch/physical/sabsoon (Last viewed 9/11/2009).
[6]C. H. Harrison and M. Siderius, "Bottom profiling by correlating beam-steered noise sequences," J. Acoust. Soc. Am. **123**, 1282–1296 (2008).
[7]P. Gerstoft, W. S. Hodgkiss, M. Siderius, C.-F. Huang, and C. H. Harrison, "Passive fathometer processing," J. Acoust. Soc. Am. **123**, 1297–1305 (2008).
[8]J. Rickett and J. Claerbout, "Acoustic daylight imaging via spectral factorization: Helioseismology and reservoir monitoring," The Leading Edge **18**, 957–960 (1999).
[9]R. L. Weaver and O. I. Lobkis, "Ultrasonics without a source: Thermal fluctuation correlations at mHz frequencies," Phys. Rev. Lett. **87**, 134301 (2001).
[10]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).
[11]P. Roux, W. A. Kuperman, and the NPAL Group, "Extracting coherent wave fronts from acoustic ambient noise in the ocean," J. Acoust. Soc. Am. **116**, 1195–2003 (2004).
[12]P. Roux, K. G. Sabra, and W. A. Kuperman, "Ambient noise cross correlation in free space: Theoretical approach," J. Acoust. Soc. Am. **117**, 79–83 (2005).
[13]K. G. Sabra, P. Roux, and W. A. Kuperman, "Arrival-time structure of the time-average ambient noise cross-correlation function in an oceanic waveguide," J. Acoust. Soc. Am. **117**, 164–174 (2005).
[14]K. G. Sabra, P. Roux, and W. A. Kuperman, "Emergence rate of the time-domain Green's function from the ambient noise cross-correlation function," J. Acoust. Soc. Am. **118**, 3524–3530 (2005).
[15]M. Siderius, "Analysis of passive seabed imaging techniques (a)," J. Acoust. Soc. Am. **123**, 3629 (2008).
[16]W. S. Burdic, *Underwater Acoustic System Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1984).
[17]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, New York, 1994).
[18]B. R. Kerman, D. L. Evans, D. R. Watts, and D. Halpern, "Wind dependence of underwater ambient noise," Boundary-Layer Meteorol. **26**, 105–113 (1983).
[19]McClelland Engineers, Inc., Ocean Bottom Survey, Air Combat Training Range, Naval Air Station Field and Laboratory Report No. 0813-0932, Brown and Root Development, Inc., Houston, TX, 1984.
[20]*APL-UW High-Frequency Environmental Acoustic Models Handbook* (Applied Physics Laboratory, University of Washington, Seattle, WA, 1994), APL-UW TR 9407, p. 128.

A stochastic response surface formulation of acoustic propagation through an uncertain ocean waveguide environment

Steven Finette
Acoustics Division, Naval Research Laboratory, Washington, DC 20375

(Received 20 February 2009; revised 21 July 2009; accepted 28 July 2009)

Stochastic basis expansions are applied to formulate and solve the problem of including uncertainty in numerical models of acoustic wave propagation within ocean waveguides. As an example, a constrained least-squares approach is used to estimate the intensity of an acoustic field whose waveguide environment has uncertainty in both source depth and sound speed. The mean intensity, a second moment of the field, and its probability distribution are computed and compared with independent Monte-Carlo computations of these quantities. Very good agreement is obtained, indicating the potential of stochastic basis expansions for describing multiple sources of uncertainty and their effect on acoustic propagation. [DOI: 10.1121/1.3212918]

PACS number(s): 43.30.Re, 43.20.Bi, 43.20.Mv [WLS] Pages: 2242–2247

I. INTRODUCTION

This paper considers a probabilistic approach for incorporating uncertainty in the modeling and simulation of acoustic propagation within ocean waveguides. The role of uncertainty in ocean acoustics has received significant attention recently due to its importance in obtaining meaningful, quantitative comparisons between measurements and model predictions[1–4] or for enhancing the predictive capability of simulation-based computations.[5,6] The emphasis here is simulation-based prediction in the presence of incomplete information concerning the system of interest; uncertainty is the term used to describe this state of affairs. This information is necessary to properly formulate the forward propagation problem and includes knowledge about the environmental parameters, fields and boundary conditions, as well as information concerning the acoustic source and receiving array. An example of the latter properties might include partial knowledge about source depth and frequency, array tilt, etc. Incomplete environmental information in ocean acoustics settings is typically associated with spatial or temporal under-sampling caused by limited resources, difficult measurement conditions, etc. Regardless of its origin, uncertainty in the context of simulation-based prediction can be viewed as a form of reducible error quite distinct from numerical truncation and discretization errors.[5] The latter two types of error are well characterized and understood, while the reducible errors, their interactions, and propagation through different system components are not well understood.

A rigorous probabilistic framework exists for representing uncertainty and is briefly reviewed in Sec. II. In this representation, stochastic basis expansions characterize incomplete information concerning both environmental and acoustic variables. At an arbitrary point in the waveguide, the method allows for the construction of a stochastic response surface[7,8] that, in the context of this study, represents the acoustic field for specified distributions of uncertain input parameters and fields. The surface is effectively a surrogate model, representing the input-output response of the system in a compact manner where the input is a particular realization of the environmental parameters and fields and the output is the pressure field. This concept has been applied to a variety of complex engineering problems, though to the author's knowledge, it has not previously been considered in the context of wave propagation problems. The response surface at an arbitrary location $\boldsymbol{r}=(r,z)$ can be computed using existing wave propagation codes, as discussed in Sec. II and illustrated by an example in Sec. III. For this approach, waveguide uncertainty is interpreted in a probabilistic sense, linked to distributions of parameters and fields that characterize the propagation over multiple realizations of the environment. Statistical information concerning the waveguide environment and source/receiver characteristics is then mapped onto the propagated acoustic field, which is now described by a stochastic (i.e., random) field. The method is based on an interpretation of polynomial chaos expansions as representations of stochastic response surfaces[7,8] in a multidimensional "uncertainty space." The chaos expansion is an example of a stochastic basis expansion. (Note that the word "chaos" refers here to the basis functionals in the expansion and is quite distinct from its use in describing extreme sensitivity to initial conditions in the context of non-linear system dynamics.) The expansion coefficients are constrained by the dynamics of acoustic propagation in the waveguide, while the expansion basis set consists of functionals of multivariate polynomials in random variables. Uncertainty is parametrized by specification of these random variables; they convey the incomplete nature of the available information through their probability density functions. Any pair of these polynomials is orthogonal with respect to a weighting function which is identified with the probability density functions of the random variables. By constructing a stochastic response surface, one has concisely summarized all information pertaining to the process. Within a wide-angle parabolic approximation to the wave equation, an example is presented that applies this method to compute statistical properties of both the mean acoustic field intensity, a second moment of the process, and the probability density function of the intensity. The basic theory and solution approach for explicitly obtaining a stochastic response surface representing these

quantities is outlined in Sec. II. An example of the method for a shallow water waveguide characterized by uncertainty in both the source depth and sound speed field is presented in Sec. III. A brief summary and conclusions are given in Sec. IV.

II. REPRESENTATION OF UNCERTAINTY THROUGH STOCHASTIC RESPONSE SURFACES

If uncertainty in the waveguide is treated probabilistically, then an acoustic field propagating through such a medium is described by a stochastic field.[5] Under these circumstances it is natural to employ stochastic basis expansions for the description of this process as well as for the uncertainty in the environment. More specifically, the acoustic field is treated here as a second-order random process. A second-order process is one with finite variance or power and represents a natural assumption for the acoustic field.

A. Polynomial chaos expansion of a random process

The stochastic representation of the acoustic field, $P(\boldsymbol{r};\boldsymbol{\xi})$, can be written as a mean-square convergent series known as a polynomial chaos expansion[9,10]

$$P(\boldsymbol{r};\boldsymbol{\xi}) = \sum_{q=0}^{\infty} \gamma_q(\boldsymbol{r})\Lambda_q(\boldsymbol{\xi}). \tag{1}$$

Vector $\boldsymbol{\xi}=(\xi_1,\ldots,\xi_D)$ in Eq. (1) represents a D-dimensional random vector that completely describes the uncertainty in the problem parametrized in terms of the random variables $\xi_1\cdots\xi_D$. Values of the elements of the vector are drawn from probability distributions and can describe uncertainty in environmental parameters as well as in source/receiver characteristics. The quantities $\gamma_q(\boldsymbol{r})$ are deterministic expansion coefficients and $\Lambda_q(\boldsymbol{\xi})$ are the chaos basis functionals. The functionals are multivariate polynomial functions[10] of the random vector with a specific form that depends on the probability density function $\rho(\boldsymbol{\xi})$. Any two polynomials satisfy an orthogonality relation with respect to the density, $\int\Lambda_q(\boldsymbol{\xi})\Lambda_{q'}(\boldsymbol{\xi})\rho(\boldsymbol{\xi})d\boldsymbol{\xi}=A\delta_{qq'}$, where A is a normalization factor depending on the choice of basis functional. Each element of $\boldsymbol{\xi}$ can be interpreted as an additional degree of freedom characterizing some component of uncertainty in the systems. Equation (1) describes the acoustic field as a random process in a factorized form where the deterministic spatial coefficients $\gamma_q(\boldsymbol{r})$ are constrained by the dynamics of wave propagation and the polynomial chaos basis functionals contain the statistical characterization of the acoustic field. The expression gives a concise representation of the statistical properties of the field due to uncertainty in both the waveguide characteristics and acoustic input specifications.

Previous application of Eq. (1) in uncertain waveguide environments involved an "intrusive" approach, in which the expansion was substituted directly into the wave equation. Applying orthogonality of the basis functionals and ensemble averaging of each resulting term then yielded a set of coupled partial differential equations for the expansion coefficients.[5,11] The intrusive approach modifies the structure of the original equation, whose solution then typically requires a new, specialized solver. Alternatively, the problem of computing the expansion coefficients is treated here in a "non-intrusive" manner.[7,8] This approach has important practical advantages over the intrusive method because (a) specialized solvers are not needed—existing propagation codes can be used to estimate the coefficients—and (b) multiple sources of uncertainty can be treated in a straightforward manner. The non-intrusive approach interprets Eq. (1) as representing, for fixed $\boldsymbol{r}$, a surface in a D-dimensional uncertainty space whose "coordinate" values are those of the coefficients of the random variables ξ_i, $i=1,\ldots,D$. In this view, Eq. (1) represents a set of stochastic response surfaces, one for each spatial location. A surface represents the set of possible acoustic field values at that location; the set of values is obtained using (probability) weighted combinations of environmental uncertainty assumed to be present in the system. The problem of describing the statistical characteristics of the acoustic field then reduces to the estimation of the expansion coefficients from a sample set of uncertain waveguide environments, each specified by a particular value of $\boldsymbol{\xi}$.

There are several approaches to estimating the coefficients in a non-intrusive formulation. In one such approach, spectral projection, Eq. (1) is multiplied by another basis functional from the set and the orthogonality relation is applied to project out the coefficients. The latter are then computed from the resulting multi-dimensional integrals using, e.g., Monte-Carlo integration. An alternative approach is applied here and involves multi-dimensional linear regression using constrained least squares. Note that the former uses the orthogonality of the basis functionals explicitly, while the latter method does not though the basis functionals are orthogonal by construction. Both the spectral projection and regression based methods rely on some kind of sampling/collocation scheme; no attempt is made here to compare these two non-intrusive approaches for estimating the coefficients.

For practical applications, the infinite series in Eq. (1) must be truncated to a finite number of terms. This is accomplished by limiting, for a given D, the maximum power of any random variable ξ that can appear in a chaos basis functional.[9] For example, a third order ($S=3$) expansion with two uncertain degrees of freedom ($D=2$) yields basis terms proportional to $\xi_1^g\xi_2^h$ for integer exponents $g,h\geq 0$, subject to the constraint that $g+h\leq 3$ for any term of the truncated series. Therefore, the highest degree of any individual variable in any term of the expansion would be either $g=3$ or $h=3$ for a third order expansion. The upper limit in Eq. (1) is given by[9,10] $T=[(S+D)!/S!D!]-1$. The choice of D and S is problem dependent. A systematic method for choosing the order of the expansion can be performed by analysis of variance methods[8] but is not pursued here. There is flexibility in the choice of the basis functionals;[10] in the example presented below, multivariate Jacobi polynomials are chosen for $\Lambda_q(\boldsymbol{\xi})$, and the corresponding probability density functions on the random variables ξ_i, $i=1,\ldots,D$, are given by beta distributions. The convergence properties of the series as a function of S have been investigated recently for wave propagation problems.[11] For computational reasons, it is important to choose the smallest values of order and dimension

for the problem of interest and methods that minimize the values of these quantities while preserving accuracy of the representation are the subject of ongoing research.

B. Construction of a stochastic response surface

The stochastic response surface at location $\boldsymbol{r}$ can be constructed by estimating the coefficients $\boldsymbol{\gamma}(\boldsymbol{r})=[\gamma_1(\boldsymbol{r})\cdots\gamma_T(\boldsymbol{r})]$ in the following manner. A representative uncertain environment is completely specified by choosing a value for each of the components of $\boldsymbol{\xi}$, where elements of $\boldsymbol{\xi}$ might represent different environmental contributions, e.g., from the water column or bathymetry as well as from acoustic properties such as source depth, etc. Choose a set of M sample environments, $\boldsymbol{\xi}^{(m)}$, $m=1,\ldots,M$, by sampling the joint density function of $\boldsymbol{\xi}$ and compute the resulting set of acoustic fields, $P^{(m)}(\boldsymbol{r};\boldsymbol{\xi}^{(m)})$, $m=1,\ldots,M$, that propagate through the set of M environments. For a given $\boldsymbol{r}$, these computations lead to a set of M linear equations for the expansion coefficients in the form

$$P^{(m)}(\boldsymbol{r};\boldsymbol{\xi}^{(m)}) = \sum_{q=0}^{[(S+D)!/S!D!]-1} \gamma_q(\boldsymbol{r})\Lambda_q(\boldsymbol{\xi}^{(m)}), \quad m=1,\ldots,M. \quad (2)$$

Note that while the equations are linear in $\gamma_q(\boldsymbol{r})$ they can be non-linear in $\boldsymbol{\xi}$, implying that the pressure field is, in general, a non-linear functional of the uncertain environment. The solution of Eq. (2) can be obtained by least-squares methods. Once the coefficients have been estimated, the response surface has been constructed and statistical properties of the field can be computed.

III. EXAMPLE OF PROPAGATION WITH MULTIPLE SOURCES OF UNCERTAINTY

Consider a far-field approximation for a harmonic acoustic field that propagates in a deterministic waveguide (i.e., no environmental uncertainty is present). The field can be written as $P(r,z)\approx\psi(r,z)[\sqrt{2/\pi k_0 r}e^{i(k_0 r-\pi/4)}]$, where $\psi(r,z)$ is the envelope function[12] and k_0 is the reference wavenumber ω/c_0. Based on the discussion in Sec. II, when the waveguide is subject to uncertainty the envelope can be treated as a random field $\psi(r,z)\rightarrow\psi(r,z;\boldsymbol{\xi})$. In the example presented below, the intensity of the acoustic field is the quantity of interest and it is proportional to $\psi^*(r,z;\boldsymbol{\xi})\psi(r,z;\boldsymbol{\xi})$. Since the intensity is also a function of $\boldsymbol{\xi}$ it can be represented as a stochastic field and written as a polynomial chaos expansion, say, $|\psi(r,z;\boldsymbol{\xi})|^2=\sum_{q=0}^{[(S+D)!/S!D!]-1}\eta_q(r,z)\Lambda_q(\boldsymbol{\xi})$, where geometrical spreading has been removed. To obtain stochastic response surfaces for $|\psi(r,z;\boldsymbol{\xi})|^2$, the sources of uncertainty must be explicitly constructed. The coefficients $\eta(r,z)$ are then estimated; note that they differ from the coefficients in Eq. (2) for the envelope.

A. Specification of the uncertainty

Two contributions of uncertainty are considered in this example, one involving the sound speed and the other involving the acoustic source. The waveguide environment is illustrated in Fig. 1. A source with a known frequency of

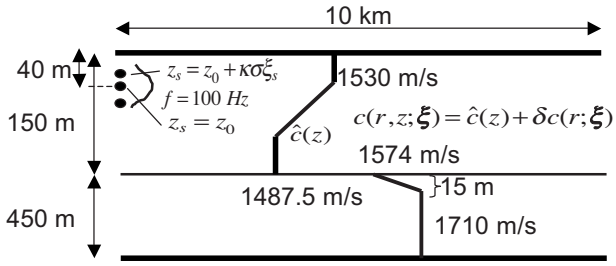


FIG. 1. Uncertain ocean waveguide with two sources of uncertainty: source depth and sound speed distribution. The assumed mean source depth is 40 m. Source frequency is 100 Hz. The uncertain sound speed $\delta c(r;\boldsymbol{\xi})$ is a zero-mean random field described by a Karhunen–Loeve expansion.

100 Hz is assumed to have an uncertain depth in a waveguide whose sediment layer lies 150 m below the surface. A single degree of freedom specified by a random variable ξ_s is sufficient to characterize the uncertainty in source depth. The acoustic source is assumed to be located in the vicinity of $z_0=40$ m (based, e.g., on *a priori* knowledge) and z_0 is interpreted here as the mean source depth. The distribution of possible depths around z_0 will be described by a beta distribution on ξ_s. An advantage of the beta distribution is its finite support in contrast, for example, to the Gaussian distribution whose long tails can be problematic since there can be a small, but non-vanishing probability of the source being located exterior to the water column. The uncertain source depth z_s can then be written as a random variable $z_s=z_0+\kappa\sigma\xi_s$, where, for generality, κ is introduced as a density-dependent parameter chosen so that σ^2 represents the variance of z_s. For a Gaussian random variable with mean zero and unit variance, $\kappa=1$, but for the beta distribution, $\kappa\neq1$ and is determined below. The general form for the (univariate) beta distribution of ξ over the support $[a,b]$ depends on two parameters α,β and is given by[10]

$$\rho_{\alpha,\beta}(\xi)=\frac{(\xi-a)^{\beta}(b-\xi)^{\alpha}}{(b-a)^{\alpha+\beta+1}B(\alpha+1,\beta+1)}, \quad a\leq\xi\leq b,\ \alpha,\beta>-1.$$

The beta function is defined as $B(u,v)=\Gamma(u)\Gamma(v)/\Gamma(u+v)$, where Γ is the factorial function. A choice of $\alpha=\beta=8$ yields a symmetric function approximating a Gaussian distribution whose finite support is limited to the interval $[a,b]$ and this choice of parameters is used in the example. A value of $\sigma=\pm4$ m is chosen for the standard deviation of z_s. To choose κ for this beta distribution one can apply the result that $\langle\xi^2\rangle=\alpha\beta/(\alpha+\beta)^2(\alpha+\beta+1)$ for a beta distribution on [0, 1]; for $\rho_{8,8}$, $\langle\xi^2\rangle=1/(4)(17)$. The support $[a=-1,\ b=1]$ is used here to construct ξ_s as a zero-mean random variable. The mapping $\xi\rightarrow2\xi-1$ shifts the distribution from $[0,1]\rightarrow[-1,1]$. The value $\kappa=\sqrt{17}$ is then obtained by noting that the shift changes the variance $\sigma^2_{2\xi-1}=2^2\sigma^2_{\xi}$. Multivariate beta distributions describing $\boldsymbol{\xi}$ can be constructed from products of the univariate distribution if the components of $\boldsymbol{\xi}$ are independent. An uncertain volumetric sound-speed field is also considered and is specified as the sum of a deterministic term $\hat{c}(z)$ representing a summer thermocline and a small, range-dependent stochastic contribution $\delta c(r;\boldsymbol{\xi})$. The stochastic term is assumed to have a spatial covariance function with the form $K(r-r')=ae^{-|r-r'|/\varepsilon}$ and a correlation length ε of 235 m. An alternative stochastic basis expansion, the

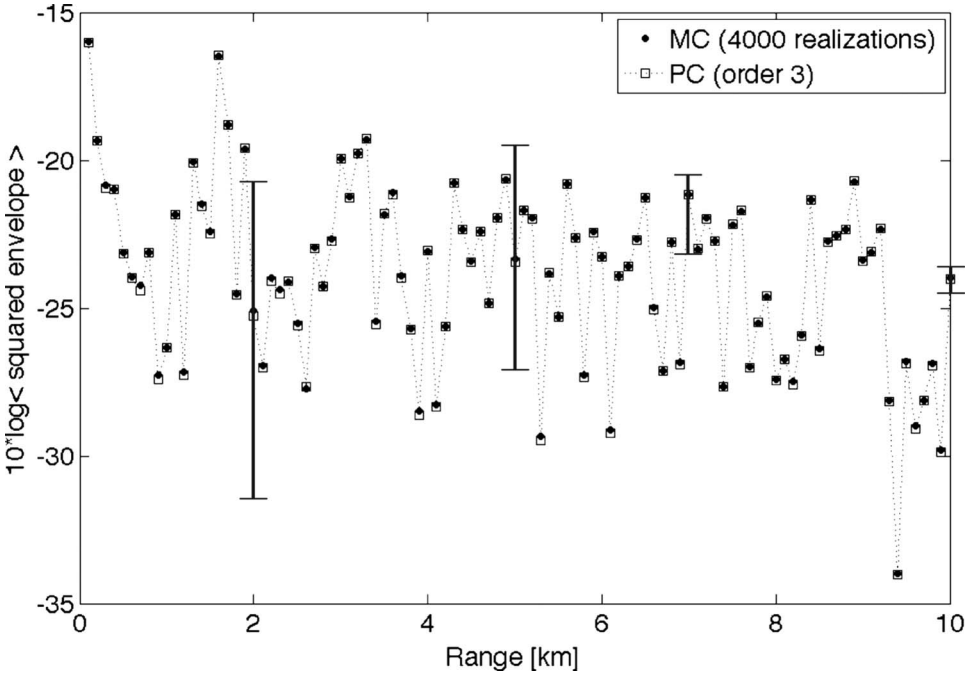


FIG. 2. Comparison between polynomial chaos and Monte-Carlo estimates of the mean intensity using a wide-angle parabolic approximation with cylindrical spreading removed. A receiver depth of 30 m is used in the computations. Vertical bars indicate intervals containing 95% of the spread in the distribution at ranges of 2, 5, 7, and 10 km from the acoustic source.

Karhunen–Loeve representation,[9] is used to express the δc contribution in sound speed through a spectral decomposition. Given the covariance function, the total sound-speed field, $c(r,z)=\hat{c}(z)+\delta c(r;\boldsymbol{\xi})$, can then be written as

$$c(r,z) = \hat{c}(z) + \sum_{i=1}^{\infty} \sqrt{\lambda_i}\phi_i(r)\xi_i, \tag{3}$$

where λ_i and ϕ_i are the eigenvalues and eigenfunctions of the covariance function. With the exponential form for K given above, these quantities can be obtained analytically.[9] In Eq. (3) the ξ_i are uncorrelated random variables and the sum is restricted to 24 terms for computational efficiency. The reason for the restriction is that the eigenvalue spectrum for an exponentially decreasing correlation function falls off rather slowly with eigenvalue number; restriction to 24 terms then yields a corresponding correlation length around 270 m for the truncated expansion. This approximation gives a reasonable estimate of the correlation function for δc, and the truncation of Eq. (3) achieves a significant dimensionality reduction in the uncertainty vector. The total uncertainty in the problem is therefore specified by a $D=25$ dimensional vector $\boldsymbol{\xi}=(\xi_1,\ldots,\xi_{24},\xi_s)$ where the first 24 components describe uncertainty in the sound-speed field, and the 25th element corresponds to the uncertainty in source depth. For simplicity, beta distributions are also used to describe the uncertainty distribution for the random variables in Eq. (3), though this is not a necessary constraint. The standard deviation of the uncertain sound speed was chosen to be ± 4 m/s. A bottom sediment layer is also present and is assumed to be known exactly. If uncertainty in the bottom was included, additional random variables would be added to $\boldsymbol{\xi}$.

B. Computation of the response surface

Since $|\psi(r,z;\boldsymbol{\xi})|^2 \geq 0$, a constrained least-squares solution for the coefficients of the chaos expansion representing $|\psi(r,z;\boldsymbol{\xi})|^2$ is obtained, with sample environments $\boldsymbol{\xi}^{(m)}$, $m=1,\ldots,4000$, drawn from beta-distributed random variables. A wide-angle parabolic equation code[13] is used to compute the sampled envelope functions for each environment.

The ensemble average, $\langle|\psi(r,z;\boldsymbol{\xi})|^2\rangle=\int|\psi(r,z;\boldsymbol{\xi})|^2 \times\rho_{8,8}(\boldsymbol{\xi})d\boldsymbol{\xi}$, is the first moment of the acoustic field intensity (with cylindrical spreading removed) or, equivalently, the second moment of the envelope itself. Note that this expression involves a D-dimensional integration over the uncertainty vector. Using the orthogonality of the basis functions, it can be shown[5] [e.g., see Eq. (31) of Ref. 5] that $\langle|\psi(r,z;\boldsymbol{\xi})|^2\rangle=\eta_0(r,z)$. This result is exact for the infinite series representation and approximate for the truncated series. Given the uncertain environment in Fig. 1, this result is compared in Fig. 2 with an independent estimate obtained from a different set of 4000 realizations of ψ computed by standard Monte-Carlo sampling techniques. A fixed receiver depth of $z=30$ m was chosen for the comparison as a function of range with range varying over 10 km. Very good agreement is achieved at each range. Given the stochastic response surfaces, probability distributions for $|\psi(r,z;\boldsymbol{\xi})|^2$ can be estimated by directly sampling the response surface for a set of environmental parameter values $\boldsymbol{\xi}'$ different from those used to construct the surface. Histograms of the distributions obtained directly from the polynomial chaos expansion are shown in Fig. 3 at several ranges (2, 5, 7, and 10 km) for a receiver depth of 30 m and compared to independent Monte-Carlo estimates of the histograms at the same ranges and depth. Note that while the initial uncertain distributions were chosen as symmetrical beta distributions, the squared envelope distributions do not generally reflect this structure, a fact due to the non-linear mapping of the waveguide uncertainty onto the acoustic field. The distributions compare very favorably at each of the four ranges with the primary exception of the values between −28 and −35 dB in Fig. 3(a). Obtained at a range of 2 km from the source, this is

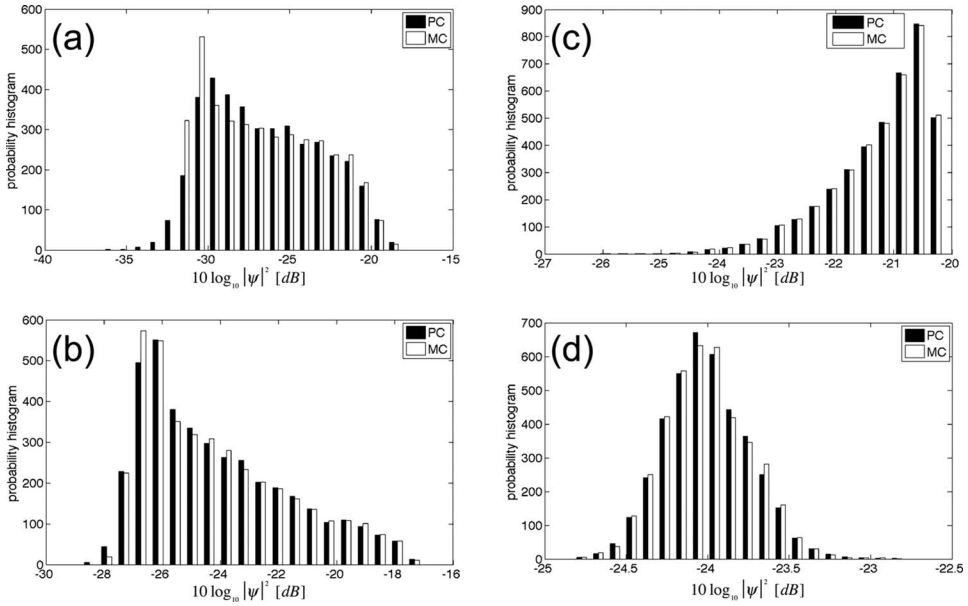


FIG. 3. Comparison of probability histograms of the mean intensity computed by polynomial chaos expansion and Monte-Carlo simulation for ranges of (a) 2 km, (b) 5 km, (c) 7 km, and (d) 10 km from the acoustic source. Receiver depth is 30 m. Note that the distributions are scaled separately.

a region of relatively higher uncertainty, as interpreted by the relatively broad nature of the distribution in comparison with those obtained at greater ranges. In addition, Fig. 2 shows percentiles at those locations with bars indicating 95% of the spread in the distributions; these are derived from the histograms for the acoustic intensity values. The spread in the distributions around the mean tends to narrow with increasing range, and this is clearly illustrated by the percentiles plotted at selected ranges in Fig. 2. A third order ($S=3$) chaos expansion was used in these computations, and the mismatch in Fig. 3(a) implies that a higher order expansion is probably necessary to capture the statistics more accurately at the shorter range. An examination of the expansion coefficients (not shown here) indicates that the uncertain source depth is the dominant source of spread in the histograms. Since one is effectively considering a distribution of depths through the random variable z_s, there also exists a corresponding distribution of modal excitation functions. Some of the higher-angle modes get stripped off at the shorter ranges and these modes are not sensed at the longer ranges, with consequently less spread (uncertainty) in the distributions of intensity at those ranges. Because only two sources of uncertainty were considered in this example (all other parameters were assumed deterministic), it is not clear that the source depth contribution will continue to dominate in a more general uncertain waveguide environment.

IV. SUMMARY AND CONCLUSIONS

Stochastic basis expansions have been applied to the problem of modeling acoustic propagation in uncertain ocean waveguide environments. The acoustic field is treated as a stochastic functional of the environment, from which statistical information can be obtained in an efficient manner. An ocean waveguide with two sources of uncertainty was considered as an example. The approach allowed for the computation of the probability distribution of the acoustic field intensity at any point from its underlying sources of uncertainty, given knowledge about the distributions of the sound speed and source depth. Mathematically, this corresponds to a solution of the uncertain (stochastic) forward problem. Very good agreement was achieved between the stochastic response surface method and independent Monte-Carlo estimates of both the intensity and its distribution over a 10 km range. While Monte-Carlo methods have traditionally been used to simulate the effects of uncertainty and were applied in this paper to obtain ground truth for comparison purposes, there is a significant need to develop alternative robust methodologies for incorporating uncertainty in complex systems.[5,9,10] One such method, involving an application of polynomial chaos expansions, has been considered here. An advantage of this approach over traditional Monte-Carlo lies in its compact representation of the random process, which can be easily manipulated to transfer uncertainty between systems. This non-intrusive representation is not simply an *ad-hoc* surface fitting of a random field with an arbitrary function but rather with a functional form linked to the distribution of environmental uncertainty. Another advantage is that by differentiating the series with respect to $\boldsymbol{\xi}$, such a representation also lends itself naturally to sensitivity analysis, where the relative importance of multiple sources of uncertainty can be assessed quantitatively. In addition, the ability to use legacy codes for computing the dynamics is a distinct advantage; problems that have traditionally been treated deterministically using complex codes developed over many years can be extended to include uncertainty without the development of special solvers.

ACKNOWLEDGMENTS

The author would like to thank Dr. Dennis Creamer for development of an algorithm to compute the chaos basis functionals as well as several useful discussions. This work was supported by funding from the Office of Naval Research under the NRL base funded program on Acoustic Field Uncertainty.

[1]C.-F. Huang, P. Gerstoft, and W. S. Hodgkiss, "Validation of statistical

estimation of transmission loss in the presence of geoacoustic inversion uncertainty," J. Acoust. Soc. Am. **120**, 1932–1941 (2006).
[2]S. Dosso and M. J. Wilmut, "Uncertainty estimation in simultaneous Bayesian tracking and environmental inversion," J. Acoust. Soc. Am. **124**, 82–97 (2008).
[3]R. A. Zingarelli, "A mode-based technique for estimating uncertainty in range-averaged transmission loss results from underwater acoustic calculations," J. Acoust. Soc. Am. **124**, EL218–EL222 (2008).
[4]P. Hursky, M. B. Porter, B. D. Cornuelle, W. S. Hodgkiss, and W. A. Kuperman, "Adjoint modeling for acoustic inversion," J. Acoust. Soc. Am. **115**, 607–619 (2004).
[5]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).
[6]K. R. James and D. R. Dowling, "A method for approximating acoustic-field-amplitude uncertainty caused by environmental uncertainties," J. Acoust. Soc. Am. **124**, 1465–1476 (2008).
[7]S. S. Isukapalli, "Uncertainty analysis of transport-transformation models," Ph.D. thesis, Rutgers University, New Brunswick, NJ (1999).
[8]S.-K. Choi, R. V. Grandhi, R. A. Canfield, and C. L. Pettit, "Polynomial chaos expansion with Latin hypercube sampling for estimating response variability," AIAA J. **42**, 1191–1198 (2004).
[9]R. Ghanem and P. Spanos, *Stochastic Finite Elements: A Spectral Approach* (Springer, New York, 2001).
[10]D. Xiu and G. Karniadakis, "The Wiener-Askey polynomial chaos for stochastic differential equations," SIAM J. Sci. Comput. (USA) **24**, 619–644 (2002).
[11]D. B. Creamer, "On using polynomial chaos for modeling uncertainty in acoustic propagation," J. Acoust. Soc. Am. **119**, 1979–1994 (2006).
[12]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP, New York, 1994).
[13]M. D. Collins, "Generalization of the split-step Pade solution," J. Acoust. Soc. Am. **96**, 382–385 (1994).

A large-aperture low-cost hydrophone array for tracking whales from small boats

B. Miller and S. Dawson
Department of Marine Science, University of Otago, P.O. Box 56, Dunedin 9054, New Zealand

(Received 25 February 2009; revised 13 August 2009; accepted 1 September 2009)

A passive sonar array designed for tracking diving sperm whales in three dimensions from a single small vessel is presented, and the advantages and limitations of operating this array from a 6 m boat are described. The system consists of four free floating buoys, each with a hydrophone, built-in recorder, and global positioning system receiver (GPS), and one vertical stereo hydrophone array deployed from the boat. Array recordings are post-processed onshore to obtain diving profiles of vocalizing sperm whales. Recordings are synchronized using a GPS timing pulse recorded onto each track. Sensitivity analysis based on hyperbolic localization methods is used to obtain probability distributions for the whale's three-dimensional location for vocalizations received by at least four hydrophones. These localizations are compared to those obtained via isodiachronic sequential bound estimation. Results from deployment of the system around a sperm whale in the Kaikoura Canyon in New Zealand are shown. [DOI: 10.1121/1.3238258]

PACS number(s): 43.30.Wi, 43.30.Sf [WWA] Pages: 2248–2256

I. INTRODUCTION

Passive acoustic localization, via arrays of hydrophones, has been used to study marine mammals for over 40 years.[1–3] Instrumentation used in previous passive acoustic studies of sperm whales has ranged from a single hydrophone,[4–7] stereo hydrophone arrays,[8–11] multihydrophone towed and vertical arrays,[12,13] and hydrophone arrays deployed simultaneously from multiple boats.[14] Recent efforts to lower the entry barriers for scientists interested in passive localization include the system presented by Hayes *et al.*[15] in 2000. This array of relatively inexpensive passive sonar buoys made mostly from commercially available off-the-shelf components was successfully used to track blue whales over several kilometers. In 2001 Møhl *et al.*[14] presented a similar unlinked sonar array, but instead of using buoys with single hydrophones, hydrophone arrays were deployed from multiple boats to track diving sperm whales.[1] By deploying a relatively deep vertical array, they were able to reconstruct three-dimensional (3D) sperm whale tracks.

Study design and hardware choice limit which software techniques are appropriate for analysis of the raw data. Beamforming techniques are most appropriately applied for short aperture towed or vertical arrays and usually yield animal locations in one or two dimensions (bearing and range). To obtain 3D animal locations, hyperbolic multilateration is usually applied to a multi-receiver large aperture array.[2,14,16–19] These hyperbolic localization techniques were originally developed for the LORAN system[20] which used radio beacons to localize ships and aircraft, and a key assumption was that the propagation speed of the transmission remained constant between the source and all receivers. The assumption of a constant speed of sound propagation is not always valid for underwater applications, especially in a place like Kaikoura, New Zealand, where several water masses converge and complex physical oceanographic processes are at work.[21]

Many of the more advanced localization algorithms involve acoustic ray-tracing and can make use of acoustic multipath detected in the recordings.[4,5,22] Ray-tracing and acoustic multipath localization algorithms require detailed knowledge of the sound velocity profile. Ray-tracing techniques can be computationally intensive especially when the hydrophone positions are not fixed with respect to each other. Additionally, multipath localization algorithms require detailed knowledge of local bathymetry.[5,23] If computation time is limited. or if detailed bathymetry and sound velocity profiles are not available, then alternative algorithms must be employed instead.

In an effort to improve on hyperbolic localization techniques, Spiesberger[24] introduced a geometric surface called an isodiachron. Isodiachronic localization is a more general form of hyperbolic localization that allows for the effective sound speed between the source and each receiver to differ. By using isodiachrons with sequential Monte Carlo methods, one can estimate not just the location of the sound source, but also the positions of the receivers and effective sound speeds. Spiesberger[25] dubbed this technique sequential bound estimation. Like many Monte Carlo methods, this technique is computationally intensive, but yields greater precision and accuracy than hyperbolic least squares error estimation under some circumstances.[26] Furthermore, sequential bound estimation does not require detailed knowledge of the speed of sound, or the bathymetry, though it can make use of this information if it is available.[25] In the absence of measured sound speed profile (SSP), it allows one to estimate a range of SSPs. Conservative estimates should be accurate (i.e., contain the true value), but may not be as precise as using a measured SSP.

In Sec. II, we present a passive sonar system that can best be described as a hybrid between the approach taken by Møhl *et al.*[14] and Hayes *et al.*[15] This system was designed for use from a single small (6 m) boat with the purpose of localizing diving sperm whales which at Kaikoura produce

loud broadband clicks at a mean click rate of 1.3 clicks/s about 60% of the time (including their time at the surface).[27] Our system consists of four free-floating buoys and one long vertical cabled array, all deployed from the same boat. Our system is based on commercially available off-the-shelf hardware, offers ease of deployment and recovery of buoys from a single platform, yet it allows measurements of three dimensional movements of nearby vocalizing whales. The system also allows measurement of bearings to whales that are many kilometers away. We describe the approach we have taken in hardware and software, and discuss the advantages and limitations of operating this array from a 6 m boat. Additionally, we show results from deployment of the system around a sperm whale in the Kaikoura Canyon in New Zealand.

II. METHODS

A. Array design

The design was heavily influenced by the limited deck space on a small vessel. The components of the array had to be compact, robust, and quickly deployed. We opted for a modular approach in order to reduce maintenance time (defective components can be quickly replaced) and to ensure that overall success in tracking did not depend too heavily on any one component. The modular approach also leaves open the possibility of adding additional instrumentation as future needs dictate.

Design specifications required sufficient battery and storage capacity to make relatively wideband recordings for several hours at a time. A further requirement common to non-linked arrays is that recordings from different platforms had to be synchronized precisely.

1. Buoys

Similar to the instrument packages of Hayes *et al.*[15] each of our buoys includes a hydrophone, a recording device, a GPS receiver, a time synchronization device, and a battery pack. Additional instrumentation included an optional fourth order bandpass filter (passband 1–40 kHz), a depth logger attached to each hydrophone, and a VHF locator beacon attached to a small mast on each of the buoys.

The recording device used in each buoy was the M-Audio Microtrack 24/96 with a 16 Gbyte compact flash card as the recording medium. Each Microtrack had stereo recording capabilities. We used a sample rate of 96 kHz (16 bit) which gave a maximum record time on 16 Gbyte media of over 11 h. While the Microtrack can record 24 data bits per sample, ambient ocean noise and electrical noise within the device itself effectively rendered this setting superfluous. Recording quality could be lowered to 44.1 kHz to enable up to 24 h of recording, though this was not attempted during this study. As compact flash cards increase in size and decrease in price, recording duration can be increased. The Microtrack recorders were powered using their internal battery which gave up to 4 h of operating time.

All hydrophones were built in-house as described by Barlow *et al.*[28] A single hydrophone was connected to each buoy with 20–30 m of shielded, harsh-environment ethernet cable. The cable chosen was TMB Proplex CAT5e which has a light Kevlar strength member allowing a maximum working pull of 140 N. Previous experience revealed that recordings made with hydrophones shallower than 20 m, resulted in increased surface noise, as well as distortion from surface echo multipaths. A 2 kg lead weight was attached to the end of the hydrophones to speed deployment, maintain hydrophone depth, and reduce hydrophone drift with respect to the buoy. Each hydrophone was connected to one channel of the recording device via a waterproof connector embedded in the buoy lid, while modulated GPS data were recorded on the other channel.

The GPS used on each buoy was the Garmin GPS-17HVS. This low cost OEM GPS was chosen because of its waterproof housing and its ability to output an accurate timing signal in addition to the raw GPS carrier phase. Via post-processing (with data from a suitable base-station), the raw carrier phase can be used to obtain highly accurate (submeter) position information. The GPS position output was connected to an FSK modulator while the GPS timing signal was connected to an amplitude modulator. Detailed description of the FSK modulation is beyond the scope of this paper, for an overview of FSK modulation with respect to this application consult Møhl *et al.*[14]

An Oceanic Veo 250 personal scuba diving computer was used to record the depth of each hydrophone throughout the duration of the deployment. The depth resolution of the dive computers was 0.3 m. Depth sensors in conjunction with post-processed GPS positions allowed for more accurate estimation of the hydrophone position. This was necessary when ocean currents and/or wind caused the hydrophone to drift so that it was not directly beneath the GPS receiver. Testing at the field site revealed this to be necessary for only the deeper boat-based array.

Each buoy had a Sirtrack VHF radio transmitter beacon mounted on a 1 m tall mast. The beacon was used to assist with relocation and recovery at the end of a recording. Beacons were tracked with a Yagi aerial connected to an Icom IC-R10 wideband scanner. Including a radio beacon on each buoy also helped with keeping track of drifting buoys during a recording. This was especially useful over large deployment areas and when sea state and weather conditions made spotting buoys difficult.

Within each buoy, the recorder, FSK circuit, bandpass filter, and two 12 V gel cell batteries were housed on a custom frame which was made from 80 mm diameter PVC drainage pipe. Slots were cut from the drainage pipe and components secured to the frame via cable ties. The frame was placed within a watertight 100 mm diameter housing also made from PVC drainage pipe. The gel cells were placed at the bottom of the frame with 1 kg of lead ballast to help the buoy maintain a vertical attitude in the water. Closed cell foam was glued around the top of the housing to provide additional floatation [Fig. 1(B)]. Buoy dimensions were 1 m in height and 100 mm in diameter. Deck space on the research vessel was limited so the four buoys were stowed upright on deck in a purpose-built wooden rack [Fig. 1(A)]. While we chose to package the instruments in a 1 m tall tube, the instruments could have fitted into an enclosure as small

FIG. 1. (A) Photograph of 4 buoys and boat-based array ready for deployment from the research vessel. (B) Buoy schematic showing arrangement of the Microtrack recorder (REC), GPS receiver, FSK modulation electronics, and power supply (BAT).

as 0.30 m long, making these buoys especially suitable for operation from small vessels. The extra space inside each buoy can be used for additional instrumentation or extra battery packs.

Hydrophones were secured to the rack next to each buoy, and hydrophone cables were wound onto the rack in a figure eight fashion to prevent tangling. By keeping buoy dimensions small and stowing them upright, deployment remained manageable even with limited deck space. A single manual switch within each buoy activated power to all electronics further facilitating speedy deployment.

2. Boat-based cabled array

The boat-based stereo array consisted of two custom built hydrophones spaced 5 m apart on 105 m of cable. The recording device used on the boat was an Edirol R-4, four channel digital recorder operating at 96 kHz sample rate with a sample resolution of 16 bits. Hydrophone and GPS data were recorded onto the first three channels, while the fourth channel was used to record dictated commentary about the situation. Commentary included descriptions of animal behaviors, vocalizations, weather information, sea state, estimated whale position, and movements of any other vessels in the area.

Using a deeper array proved crucial for obtaining accurate 3D localizations of the target animal. The short distance between hydrophones of the boat-based array also facilitated tracking individual animals when several were vocalizing at the same time. The boat-based array functioned as a short-aperture vertical array which was used to measure the vertical bearing to vocalizing animals using a custom MATLAB script. For short time scales (tens of seconds), vocalizations with widely different bearings are likely to come from different individuals. Similarly, bearings coming from the same individual would be expected to change gradually over short time scales. These assumptions were used by the classification algorithm to reduce ambiguities during analysis that occurred due to multiple vocalizing animals.

In addition to the recordings made using the stereo hydrophone array, a handheld directional hydrophone and compass were used to estimate the range and bearing to the target animal(s) throughout the deployment. Tracking the whale this way provided feedback necessary to reposition hydrophones that drifted too far from the target whale. Because the depth resolution of our system was provided primarily from the boat-based array, it was especially important that the boat-based array remain close enough to detect the animal continuously throughout the recording. In addition to constant feedback, the range and bearing estimate also provided an independent check on the validity of localizations obtained from the non-linked array.

B. Deployment and recovery

Before deploying the buoys and boat-based array, sperm whales were tracked using a custom-built directional hydrophone. Whales were tracked until they surfaced in order to obtain an identification photograph[29,30] before deploying the array. The photographed whale (target whale) was typically tracked via directional hydrophone for 20–25 min to ascertain its speed and heading (if any) before deploying the first buoy. Because sperm whales in Kaikoura can travel several kilometers from fluke up, to fluke up,[31] it was important to have an estimate of the target whale's position and general direction of movement before deploying the buoys.

Buoys were typically deployed in either a triangle or square configuration surrounding the animal(s) of interest, with the boat and stereo hydrophone array (and with a bit of luck the target whale) at the center of the polygon. Typical deployment distances between adjacent buoys were 1.5–2 km. To deploy four buoys typically took 20 min; however, poor weather, other vessels, and navigational hazards such as fishing gear increased deployment time.

To deploy buoys, the boat engine was stopped, the buoy electronics turned on, and the Microtrack set to record. The hydrophone cable was deployed before placing the buoy overboard, and the level meter on the Microtrack was checked to make sure that both the hydrophone and GPS recording chains were functioning and that the recorder gains were set appropriately The waterproof lid was replaced, and the VHF locator beacon was attached before placing the buoy over the side of the boat. To facilitate recovery, each buoy's deployment location was marked on the vessel's navigation GPS.

Some care was required when setting the recording gain for each buoy because this could not be changed after deployment. Over the course of the recording, whale(s) can swim toward some buoys and away from others, causing a large change in received level. The dynamic range of the Microtrack recording unit was not wide enough to accommodate this change in received level without gain adjustment. Had the 24-bit recordings on the Microtrack not been dominated by electrical noise, then changing the bit resolution from 16 to 24 bits would have greatly helped handle this limitation. While hardware-based automatic gain control might have been a possible solution, we did not pursue this. Because the goal of this study was localization of whales, we opted for a higher gain setting in order to detect whales further away and maximize buoy separation. This choice came at the expense of clipping some loud whale vocalizations.

When the target animal was unlikely to be audible on a minimum of four hydrophones, buoys could be repositioned; however, this required an interruption in monitoring with the

boat based array which effectively limited localization accuracy to two dimensions (x and y) during this time. While buoys could be repositioned by a dedicated support vessel, this would effectively double the operating costs of the array and has not been attempted. More often instead of repositioning buoys the recording was terminated and the buoys recovered. After recovery, data were downloaded to a personal computer (PC) for synchronization and analysis.

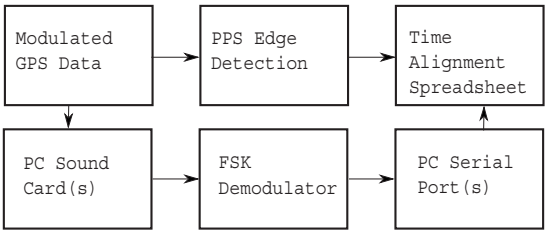


FIG. 2. Data flow for time alignment and demodulation of recorded GPS data.

C. Synchronization

The standard procedure for localizing animals with an acoustic array involves computing time of arrival differences (TOADs) between each pair of hydrophones for each vocalization.[2,3,14,15,17,18,32–36] When using an unlinked array, all recordings made at each location must be synchronized before TOADs can be computed accurately. Synchronization must address both jitter and clock drift. For the purposes of this article, jitter can be thought of as very short-term changes in the sample rate, while clock drift refers to long-term differences between the device's clock and the GPS synchronization signal. Both of these errors arise from imperfections in the digital clock used for analog-to-digital conversion in the recording unit.

For our system, jitter, measured during synchronization, was typically on the order of 0.002% for all devices. Clock drift was also measured and was typically between 0.5 and 2 ms/min for all devices. Measurement of the jitter and drift rates is not only necessary for accurate localization but also provides a measure of the temporal fidelity of the audio device. Audio time alignment and jitter/drift correction was performed via a two-stage process. The first step involved coarse alignment, which synchronized the start and end of each recording to within 1 s and assumes constant drift and no jitter. The second step (fine scale alignment) provided sample-accurate audio synchronization once every second for the duration of the recorded audio.

The GPS position information stored in the FSK-modulated audio signal included latitude, longitude, as well as UTC date and time of the signal with time resolution of 1 s. For coarse alignment, we extracted this information from the first and last seconds of the recording to compute the GPS start and end seconds for the recording. Subtracting the ending GPS second from the starting second yields the GPS duration, t_{GPS}. The average clock drift rate was computed as $(t_{\mathrm{GPS}}-t_{\mathrm{recording}})/t_{\mathrm{GPS}}$, where $t_{\mathrm{recording}}$ was the total number of audio samples per channel divided by the nominal sample rate (96 kHz). This coarse alignment does not account for jitter or inaccuracies resulting from a non-constant clock drift rate over the duration of the recording. To investigate these errors and account for them if they are present, we used the Garmin GPS 17 timing signal, which is a 1 Hz pulse wave with a duty cycle of 0.1. The rising edge of this pulse marked the start of the GPS second with a nominal accuracy of ± 1 μs (Garmin GPS 16/17 Technical manual). When the 100 ms pulse was active, it reduced the signal amplitude of FSK-modulated GPS data. The instantaneous sample rate of the recording unit was computed by simply counting the number of samples between successive pulse edges, and the instantaneous jitter was computed as the difference between nominal sample rate and instantaneous sample rate.

Custom synchronization software written using Mathworks MATLAB was used to detect the sample number corresponding to the leading edge of the amplitude-modulated timing pulse. This edge detection software began by loading 1 s of audio into memory. This audio was divided into ten consecutive sequences and the rms amplitude of each sequence was computed. Due to the amplitude modulation, the sequences containing the timing pulse had a different rms amplitude than the rest of the signal. The earliest sequence with a different rms amplitude contained the leading edge of the pulse. This sequence was kept, while the others were discarded. For each sample in the remaining sequence, the rms value of the subsequent 20 samples was computed. The difference between sequential rms values was computed and the sample with the largest change in rms amplitude corresponded to the leading edge of the PPS. This process was repeated for the duration of the recorded audio.

Simultaneous to edge detection, the FSK-modulated audio track was played into a custom-made hardware demodulator and the GPS data were recorded via a PC serial port. Hardware FSK demodulation with concurrent software PPS detection allowed for synchronization of multiple buoys at the same time. The audio sample number of the PPS edge, latitude, longitude, UTC time, and raw carrier phase information from each platform was written to a synchronization data file for every second of audio processed (Fig. 2). By using a hardware-based demodulator circuit for each of the five audio channels, 5 channel-hours of modulated GPS position and timing information could be decoded in 1 h. This proved to be significantly faster than our best attempts at implementing software-based demodulation as described by Møhl *et al.*[14] and can work in real time provided that there are as many demodulators and serial ports as there are modulated GPS signals.

Hydrophone depth sensors were activated via a water contact switch. Depth data were synchronized with the GPS data using either the audible tone made when the depth sensor was active for a predetermined amount of time or the sound made from the entry of the depth sensor into the water, both of which were audible in the recording for each hydrophone. At the end of preprocessing, the multichannel recordings of the whales' sounds, location data for each buoy and the boat, and the depth data for each hydrophone were synchronized.

D. Detection and localization

Recently there have been many different techniques proposed for detection, classification, and localization of sperm whale clicks. To obtain 3D whale positions, we implemented a selection of detection, classification, and localization algorithms[22,25,37,38] using Mathworks MATLAB 7.3, and adapting the methods for use with our system as necessary.

1. Detection

For detection of sperm whale vocalizations, audio recordings from each platform were bandpass filtered between 2 and 20 kHz, a band which contains most of the energy of typical sperm whale vocalizations. Vocalizations were detected from filtered recordings using Page's test, which is an energy detector.[37] Specifically, we followed the algorithm outlined in Ref. 39, Part II, Sec. I. While there have recently been numerous methods for detection and classification of sperm whale clicks,[40–44] Page's test was chosen because its implementation was intuitive, fast, and it has been used successfully in previous studies involving localization of sperm whales.[39,45] Detection parameters that yielded good agreement with visual inspection of the spectrogram for the first few minutes of audio were selected for use. The detection threshold, V_1, was set to 16 (24 dB); the end of detection threshold, V_0, was 1; and the exponential weighting on the noise, α, was set to 0.9 (notation follows Ref. 39). Automating the detection process was necessary for analyzing the large number of recorded sperm whale vocalizations.

2. Bearing localization

TOADs were computed between both hydrophones from the boat-based stereo array using cross-correlation of the waveform of detected clicks. The time lag at the peak in the cross-correlation function was recorded as the TOAD of a direct arrival. Because the distance between these hydrophones was much smaller than the distance to the target whale, these TOADs provide a measure of the angle of arrival of the sound. These angles were plotted as a function of time to yield a bearing-time plot for the recording. Bearing tracks were traced by a human operator and traces were numbered and assigned to an individual whale via a custom MATLAB interface. Bearings were traced with the following criteria. Bearings that corresponded to an individual whale track must change slowly and continuously over time. This constraint eliminated noise sources from being selected as a whale trace. Any ambiguities in a trace, such as the intersection of multiple traces or gaps longer than 7 min, resulted in the termination of a trace and the start of a new trace at a time after the ambiguities could be resolved. A recording typically contained between 1 and 6 individual bearing traces at any one time.

3. Surface echo detection

Echo detection based on autocorrelation was performed on vocalizations from each bearing trace from the vertical array. For each vocalization, the absolute value of the autocorrelation of the waveform was computed. The largest peak in this autocorrelation function that occurred between 10 and 200 ms after the direct arrival was considered a surface echo so long as the time lag of this peak did not correspond to a direct arrival from another bearing trace. All surface reflections from a particular bearing trace were written to a separate log file. These surface reflections can be thought of as arriving at a virtual hydrophone that mirrors the real hydrophone above the ocean surface.[5,9,46] These virtual hydrophones were used as additional receivers and increase both the number of hydrophones in the array and the vertical hydrophone separation, thus increasing the localization performance of the array.[16]

4. Classification (click association)

The inter-click interval from each of the bearing traces was computed and used as input into a custom MATLAB program that implemented the "rhythm analysis" algorithm described by Thode.[22] This algorithm was necessary to associate vocalizations received at each buoy and virtual hydrophone with vocalizations received from an individual whale at the stereo array. When a vocalization was matched at 4 or more real hydrophones, arrival time differences were calculated between all hydrophone pairs by computing the cross-correlation of the audio for each matching detection. The time lags of the peak of the cross-correlation function were stored as TOADs and used in further localization analysis.

5. Localization

Once all TOADs have been computed, these data as well as the hydrophone positions were used as input into a MATLAB program that implemented the hyperbolic localization algorithm described by Spiesberger and Fristrup.[38] To estimate the localization precision, a separate sensitivity analysis was performed.

For the sensitivity analysis, we assigned uncertainty to each of the model inputs and created uniform probability distributions based on the measured data and estimated/measured uncertainty for each of the model inputs. Variance for the horizontal hydrophone position was measured to be ± 2 m based on a 48 h comparison of each GPS receiver to a surveyed reference position. Variance for each hydrophone depth sensor was assumed to be ± 0.3 m according to the manufacturer's specifications. Effective sound speeds were allowed to vary across the range of sound speeds computed with equations from Del Grosso[47] using historical monthly temperature, salinity, and depth data for the study area from the World Ocean Atlas.[48,49] TOAD variance was computed according to Spiesberger and Fristrup[38] equation 41. We then drew 2000 samples from each of these random variables and used each set of samples as input to the localization algorithm to obtain a cloud of points that represents the whale's position.

For the x, y, and z coordinates of the whales position, probability distributions, P_x, P_y, and P_z, were estimated at each time step from the output of the sensitivity analysis. Estimates of P_x, P_y, and P_z were calculated from the normalized histograms of each x, y, and z coordinate of the whale's position using bin widths of 1 m. The total volume for each

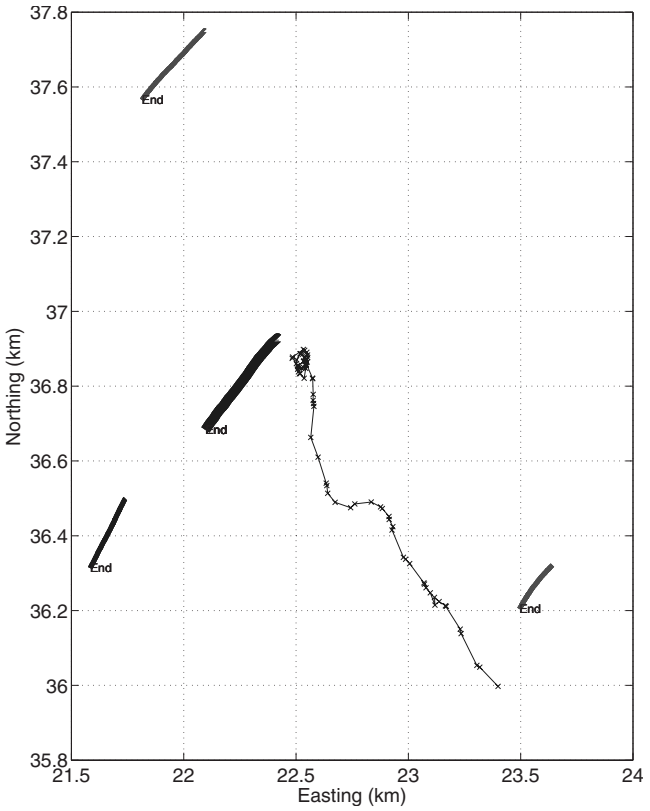


FIG. 3. October 30, 2007 array deployment geometry. Bold line shows the track of the boat, while normal lines show the track of the buoys. The thin line with crosses is the whale track, which shown in detail in Fig. 4.

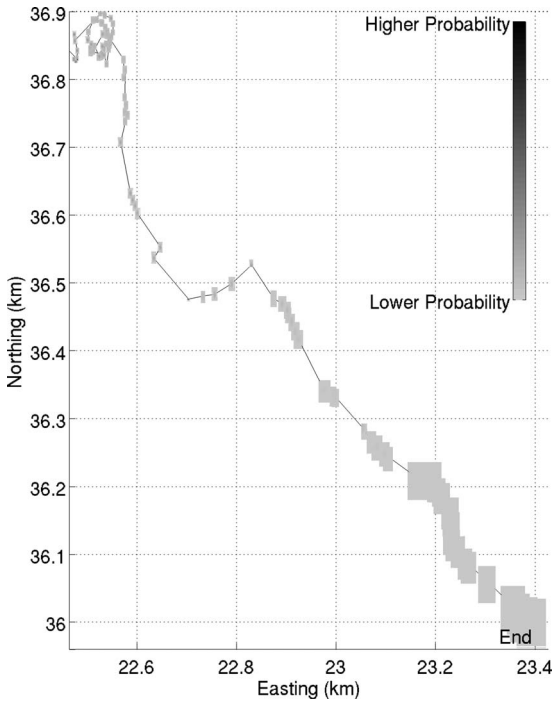


FIG. 4. Joint *X-Y* whale position probability from isodiachronic sequential Monte Carlo analysis. The whale circles as he dives initially and then heads from northwest to southeast.

localization cloud was computed as $(\hat{P}_x - \check{P}_x)(\hat{P}_y - \check{P}_y)(\hat{P}_z - \check{P}_z)$, where $\hat{P}$ and $\check{P}$ denote the maximum and minimum values from the probability distributions. The total cloud volume is a measure of localization precision. A threshold volume of 1.77×10^6 m^3, which is equal to the volume of a sphere with a diameter of 150 m (approximately 10 whale lengths), was used to exclude localizations with low precision (Fig. 6).

Because hyperbolic localization can yield incorrect results in a stratified environment, isodiachronic sequential bound estimation[25] was used to spot-check the whale's position at 15 s intervals starting from the first vocalization. While isodiachronic sequential bound estimation is more accurate than hyperbolic localization, our implementation of this method was computationally intensive and would have taken prohibitively long to analyze every vocalization this way. Performing the sequential bound localization every 15 s served as a quality control check on the hyperbolic localization results.

The same random variables created for the sensitivity analysis were used as inputs into the isodiachronic sequential bound localization algorithm to obtain a cloud of potential whale positions.[25] The shape of this cloud reflects the optimal localization precision and accuracy of the system without requiring the constraint of a homogeneous environment. When localization clouds from the sequential bound estimation are drastically different than those from the hyperbolic localization algorithm, then the assumption of an isovelocity sound speed is likely to be invalid.[26]

E. Trial deployment

On 30 October 2007, the array was deployed around a single male sperm whale diving over the Kaikoura canyon (Fig. 3). The array was deployed from the research vessel Grampus, a 6 m aluminum boat, operating over the Kaikoura canyon with a crew of two. Using the directional hydrophone, three sperm whales were detected vocalizing; however, only one sperm whale was estimated to be within the bounds of the array at the time of deployment. When possible the range and bearing to the diving whale were measured at the surface using a hand bearing compass and laser range finder (Bushnell Yardage Pro Compact 600).

III. RESULTS

Directional hydrophone estimates of the whale's position indicate that the whale dived near the research vessel and moved toward the southeast. This is consistent with the track generated by the isodiachronic sequential bound estimate (Fig. 4). However, estimates from the directional hydrophone were not precise enough to reveal that the whale descended along the path of a spiral, which can be seen in the 3D tracks computed via sequential bound estimation (Fig. 5). The precision of localization decreased as the whale moved away from the center of the array (Figs. 6 and 7). When localization algorithms could make use of surface reflected multipath the localization precision substantially increased (Fig. 6), which is consistent with results described by Wahlberg *et al.*[16]

The median whale position from the marginal distributions of the hyperbolic sensitivity analysis, P_x, P_y, and P_z, fell within the 95% confidence intervals from the isodiachronic sequential bound analysis (Fig. 7). Maximum depth

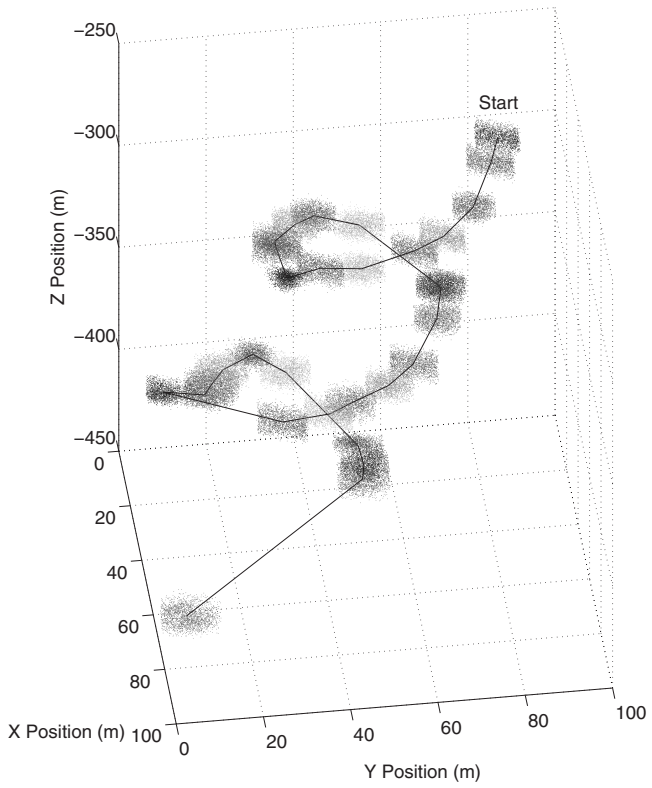


FIG. 5. Whale trajectory in 3D from the beginning of the dive (time 00:26:15–00:35:32). Whale spirals as he descends. Each cloud corresponds to one whale vocalization. For clarity successive clouds have different shadings and the solid line connects the median point from each cloud.

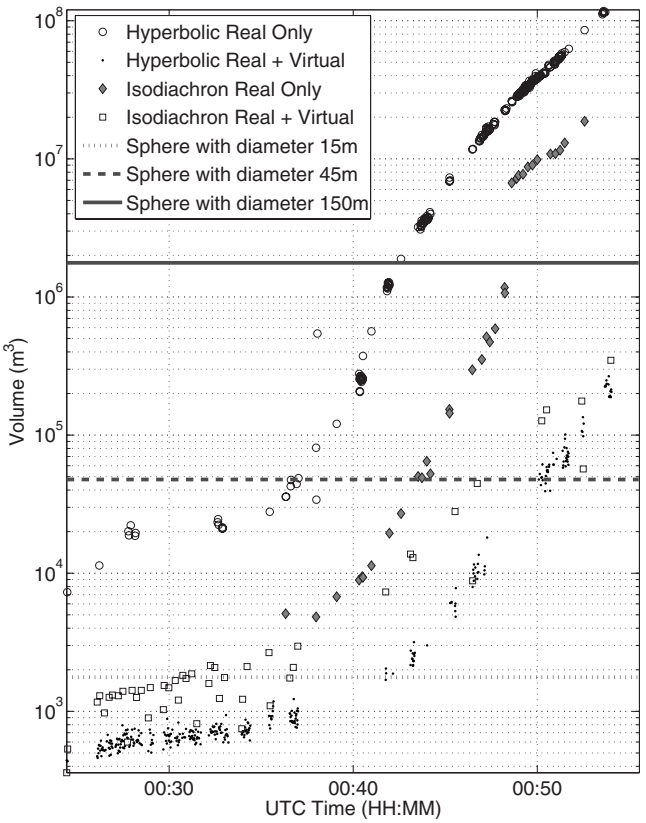


FIG. 6. The total volume of each localization cloud. Open circles and filled circles show volumes from hyperbolic sensitivity analysis using real hydrophones and using virtual hydrophones from multipath surface reflections, respectively. Squares and diamonds show isodiachron sequential bound analysis again using real and virtual hydrophones. Solid line indicates the precision threshold. Dashed and dot-dashed lines show reference volumes corresponding to spheres with diameters of 45 and 15 m (3 and 1 whale-length), respectively.

was 599 m, while mean depth was 418 m which is comparable to sperm whale diving depths measured in other parts of the world.[34,50]

IV. DISCUSSION

The array has been successfully used to localize diving sperm whales in the Kaikoura canyon. Results and error estimates obtained are consistent with those obtained from other passive sonar systems used to track sperm whales in 3D.[17,19,34,45] Tracking whales in three dimensions has the potential to yield information that may not be observable from tagging animals with a depth logger. The spiral at the beginning of the dive in Fig. 5 is a clear illustration of an advantage of 3D tracking. As computer processors, analog-to-digital converters, and digital storage become more powerful and affordable, so should the ability to create inexpensive passive sonar systems with higher fidelity hardware and more sophisticated on-board software.

As computing power increases, it should become feasible to perform the more accurate isodiachronic analysis for every vocalization instead of using it primarily as a quality control check. By using the fast hyperbolic analysis method to localize every click and validating the results with the slower isodiachronic analysis, we attempt to strike a compromise between computation time and accuracy. By using confidence intervals from the sequential bound analysis instead of the hyperbolic analysis, we effectively trade the higher temporal resolution of the hyperbolic analysis for higher overall accuracy of the isodiachronic analysis. This trade-off is only necessary when computing power or analysis time are limited.

Our system does have a few additional limitations. A notable limitation is the limited dynamic range of the recording units, resulting from our choice of inexpensive off-the-shelf field recorders. Because the goal of our system was 3D localization rather than measurement of sonar emission patterns, we view this trade-off as acceptable since it allows detection and localization over greater ranges. Automatic gain control or recording devices with wider dynamic range would address this issue. Another limitation of the system is the amount of time required to process the data from each platform to obtain localization. Presently the largest portion of processing time is spent demodulating the GPS positions, which takes the same amount of time as the duration of the recording. Lastly, a major limitation of the system is that high accuracy localizations only occur when the whale is within the bounds of the array. For whales that can swim several kilometers in a single dive, some luck is required to obtain high accuracy localizations over a full dive cycle. However, even the subset of observations for which the whale remains inside the bounds of the array has the potential to yield important insights into the underwater behavior of these whales.

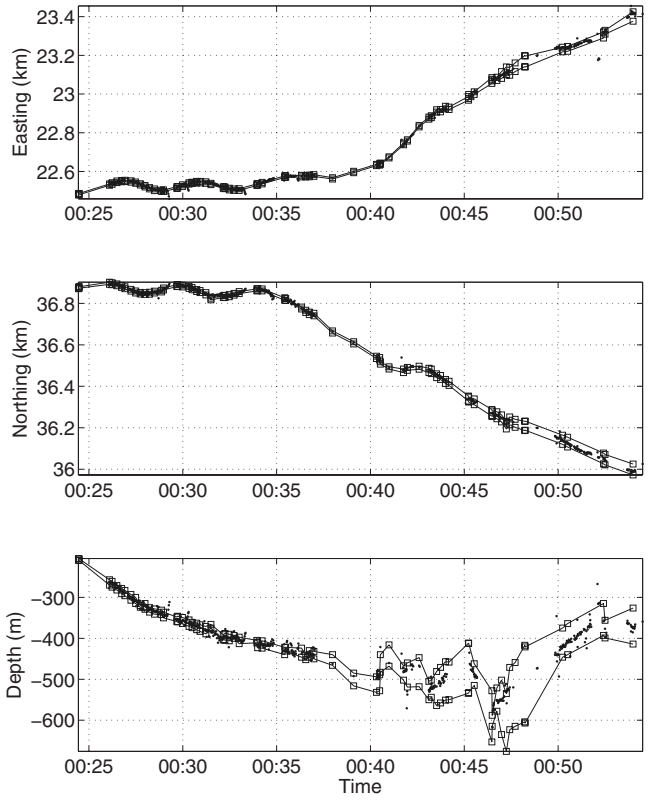


FIG. 7. Whale location as a function of time. Dots show the median of the marginal probability distribution from the hyperbolic sensitivity analysis. Squares show the 95% confidence limits from the isodiachronic sequential bound estimation.

The principal advantages of our system are its low cost, portability, and ease of use from a small boat. The instrumentation has no moving parts and can survive bumps and jostles that occur at sea during difficult weather conditions. The system is portable and unlike a fixed hydrophone array, it can be deployed and repositioned around the target animals. Processing occurs on shore and requires a desktop PC with adequate storage space (5–10 Gbytes/recording session). Each one of our buoys can be built using mostly off-the-shelf components with a total cost of the parts under US$1000.

While the main goal of our study is to detect and localize sperm whales in the Kaikoura canyon, the array could potentially be used to localize any loud sound sources in the area including baleen whale vocalizations, shipping traffic, underwater explosions, or construction activity. On one occasion the system was used to localize concurrently not only a nearby diving sperm whale but also a singing humpback whale in Kaikoura at an approximate range of 8 km (B. Miller, unpublished).

Long-term use of the system has the potential to provide insight into whether individual whales have different foraging styles and how diving behavior changes with season. Additionally the system may be used to investigate the effects of anthropogenic noise from sources such as whale watching platforms on sperm whale underwater behavior.

ACKNOWLEDGMENTS

Thanks to Miranda van de Linde, Elanor Hutchison, Liz Slooten, and all of our volunteers for providing help with field work. We thank Hamish Bowman for help with buoy construction and design advice, Ross Vennell for assistance writing localization software, and Paul Denys for many consultations about GPS postprocessing. Bertel Møhl and Niels Kristiansen generously provided their FSK modulator and demodulator designs, as did Aleks Zosuls for bandpass filter designs. Funding for field work was provided by the New Zealand Whale and Dolphin Trust. Funding for instrumentation was provided by Otago University.

[1]W. Cummings, B. Brahy, and W. Herrnkind, "The occurrence of underwater sounds of biological origin off the west coast of Bimini, Bahamas," in *Marine Bio-Acoustics: Proceedings of a Symposium Held at the Lerner Marine Laboratory, Bimini, Bahamas*, edited by W. Tavolga, Pergamon, New York 1964, pp. 27–43.
[2]W. Watkins and W. Schevill, "Sound source location by arrival-times on a non-rigid three-dimensional hydrophone array," Deep-Sea Res. **19**, 691–706 (1972).
[3]C. W. Clark, W. T. Ellison, and K. Beeman, "Acoustic tracking of migrating bowhead whales," *Proc. IEEE Oceans '86* (IEEE, New York, 1986), pp. 341–346.
[4]C. Laplanche, O. Adam, M. Lopatka, and J.-F. Motsch, "Male sperm whale acoustic behavior observed from multipaths at a single hydrophone," J. Acoust. Soc. Am. **118**, 2677–2687 (2005).
[5]C. O. Tiemann, A. M. Thode, J. Straley, V. O'Connell, and K. Folkert, "Three-dimensional localization of sperm whales using a single hydrophone," J. Acoust. Soc. Am. **120**, 2355–2365 (2006).
[6]M. Q. Rhinelander and S. M. Dawson, "Measuring sperm whales from their clicks: Stability of interpulse intervals and validation that they indicate whale length," J. Acoust. Soc. Am. **115**, 1826–1831 (2004).
[7]N. Jaquet, S. Dawson, and L. Douglas, "Vocal behavior of male sperm whales: Why do they click?," J. Acoust. Soc. Am. **109**, 2254–2259 (2001).
[8]J. Barlow and B. Taylor, "Estimates of sperm whale abundance in the Northeastern temperate Pacific from a combined acoustic and visual survey," Marine Mammal Sci. **21**, 429–445 (2005).
[9]E. K. Skarsoulis and M. A. Kalogerakis, "Ray-theoretic localization of an impulsive source in a stratified ocean using two hydrophones," J. Acoust. Soc. Am. **118**, 2934–2943 (2005).
[10]A. Thode, "Three-dimensional passive acoustic tracking of sperm whales (*Physeter macrocephalus*) in ray-refracting environments," J. Acoust. Soc. Am. **118**, 3575–3584 (2005).
[11]B. K. Nielsen and B. Møhl, "Hull-mounted hydrophones for passive acoustic detection and tracking of sperm whales (*Physeter macrocephalus*)," Appl. Acoust. **67**, 1175–1186 (2006).
[12]V. Teloni, "Patterns of sound production in diving sperm whales in the Northwestern Mediterranean," Marine Mammal Sci. **21**, 446–457 (2005).
[13]A. Heerfordt, B. Møhl, and M. Wahlberg, "A wideband connection to sperm whales: A fiber-optic, deep-sea hydrophone array," Deep-Sea Res., Part I **54**, 428–436 (2007).
[14]B. Møhl, M. Wahlberg, and A. Heerfordt, "A large-aperture array of nonlinked receivers for acoustic positioning of biological sound sources," J. Acoust. Soc. Am. **109**, 434–437 (2001).
[15]S. A. Hayes, D. K. Mellinger, D. A. Croll, D. P. Costa, and J. F. Borsani, "An inexpensive passive acoustic system for recording and localizing wild animal sounds," J. Acoust. Soc. Am. **107**, 3552–3555 (2000).
[16]M. Wahlberg, B. Møhl, and P. T. Madsen, "Estimating source position accuracy of a large-aperture hydrophone array for bioacoustics," J. Acoust. Soc. Am. **109**, 397–406 (2001).
[17]P. Giraudet and H. Glotin, "Real-time 3d tracking of whales by echo-robust precise TDOA estimates with a widely-spaced hydrophone array," Appl. Acoust. **67**, 1106–1117 (2006).
[18]R. Morrissey, J. Ward, N. DiMarzio, S. Jarvis, and D. Moretti, "Passive acoustic detection and localization of sperm whales (*Physeter macrocephalus*) in the tongue of the ocean," Appl. Acoust. **67**, 1091–1105 (2006).
[19]L. F. Baumgartner M. F., J. Partan, K. Ball, and K. Prada, "Tracking large marine predators in three dimensions: The real-time acoustic tracking system," IEEE Transl. J. Magn. Jpn. **33**, 146–157 (2008).
[20]J. P. Van Etten, "Loran c system and product development (loran c principles, ground station operation and navigation equipment)," Electrical Communication **45**, 100–115 (1970).
[21]F. Lariviere, "Sperm whale habitat off Kaikoura, New Zealand, a descriptive study of physical and biological processes influencing sperm whale

distribution," MS thesis, Otago University, Dunedin, New Zealand (2001).
[22]A. Thode, "Tracking sperm whale (*Physeter macrocephalus*) dive profiles using a towed passive acoustic array," J. Acoust. Soc. Am. **116**, 245–253 (2004).
[23]C. Tiemann, A. Thode, J. Straley, K. Folkert, and V. O'Connell, "Model-based passive acoustic tracking of sperm whale foraging behavior in the Gulf of Alaska," J. Acoust. Soc. Am. **118**, 1909–1909 (2005).
[24]J. L. Spiesberger, "Geometry of locating sounds from differences in travel time: Isodiachrons," J. Acoust. Soc. Am. **116**, 3168–3177 (2004).
[25]J. L. Spiesberger, "Probability distributions for locations of calling animals, receivers, sound speeds, winds, and data from travel time differences," J. Acoust. Soc. Am. **118**, 1790–1800 (2005).
[26]J. Spiesberger and M. Wahlberg, "Probability density functions for hyperbolic and isodiachronic locations," J. Acoust. Soc. Am. **112**, 3046–3052 (2002).
[27]L. A. Douglas, S. M. Dawson, and N. Jaquet, "Click rates and silences of sperm whales at Kaikoura, New Zealand," J. Acoust. Soc. Am. **118**, 523–529 (2005).
[28]J. Barlow, S. Rankin, and S. Dawson, "A guide to constructing hydrophones and hydrophone arrays for monitoring marine mammal vocalizations," NOAA Technical Memorandum NMFS, Technical Report No. 417, National Oceanic and Atmospheric Administration, Southwest Fisheries Science Center, La Jolla, CA, 2008.
[29]T. Arnbom, "Individual identification of sperm whales," Rep. Int. Whal. Comm. **37**, 201–204 (1987).
[30]S. Childerhouse, S. Dawson, and E. Slooten, "Abundance and seasonal residence of sperm whales at Kaikoura, New Zealand," Can. J. Zool. **73**, 723–731 (1995).
[31]N. Jaquet, S. Dawson, and E. Slooten, "Seasonal distribution and diving behaviour of male sperm whales off Kaikoura: Foraging implications," Can. J. Zool. **78**, 407–419 (2000).
[32]L. E. Freitag and P. L. Tyack, "Passive acoustic localization of the Atlantic bottlenose dolphin using whistles and echolocation clicks." J. Acoust. Soc. Am. **93**, 2197–2205 (1993).
[33]J. L. Spiesberger, "Locating animals from their sounds and tomography of the atmosphere: Experimental demonstration," J. Acoust. Soc. Am. **106**, 837–846 (1999).
[34]M. Wahlberg, "The acoustic behaviour of diving sperm whales observed with a hydrophone array," J. Exp. Mar. Biol. Ecol. **281**, 53–62 (2002).
[35]W. Zimmer, M. Johnson, A. D'Amico, and P. Tyack, "Combining data from a multisensor tag and passive sonar to determine the diving behavior of a sperm whale (*Physeter macrocephalus*)," IEEE J. Ocean. Eng. **28**, 13–28 (2003).
[36]E.-M. Nosal and N. L. Frazer, "Track of a sperm whale from delays between direct and surface-reflected clicks," Appl. Acoust. **67**, 1187–1201 (2006).
[37]D. Abraham, "Passive acoustic detection of marine mammals using Page's test," *Applied Sequential Methodologies: Real-World Examples With Data Analysis*, edited by N. Mukhopadhyay, S. Datta, and S. Chattopadhyay (Marcel Dekker, New York, 2004).
[38]J. Spiesberger and K. Fristrup, "Passive localization of calling animals and sensing of their acoustic environment using acoustic tomography," Am. Nat. **135**, 107—153 (1990).
[39]W. M. X. Zimmer, P. L. Tyack, M. P. Johnson, and P. T. Madsen, "Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis," J. Acoust. Soc. Am. **117**, 1473–1485 (2005).
[40]O. Adam, "Advantages of the Hilbert Huang transform for marine mammals signals analysis." J. Acoust. Soc. Am. **120**, 2965–2973 (2006).
[41]O. Adam, "The use of the Hilbert-Huang transform to analyze transient signals emitted by sperm whales," Appl. Acoust. **67**, 1134–1143 (2006).
[42]V. Kandia and Y. Stylianou, "Detection of sperm whale clicks based on the Teager-Kaiser energy operator," Appl. Acoust. **67**, 1144–1163 (2006).
[43]J. P. Larue, G. E. Ioup, and J. W. Ioup, "Detecting sperm whale clicks in the presence of ambient and shipping noise using higher order moments," J. Acoust. Soc. Am. **115**, 2487 (2004).
[44]M. Lopatka, O. Adam, C. Laplanche, J.-F. Motsch, and J. Zarzycki, "Sperm whale click analysis using a recursive time-variant lattice filter," Appl. Acoust. **67**, 1118–1133 (2006).
[45]E. Nosal and L. Frazer, "Sperm whale three-dimensional track, swim orientation, beam pattern, and click levels observed on bottom-mounted hydrophones," J. Acoust. Soc. Am. **122**, 1969 (2007).
[46]A. M. Thode, "The use of acoustic multipath for localization of sperm whales," J. Acoust. Soc. Am. **116**, 2606 (2004).
[47]V. A. Del Grosso, "New equation for the speed of sound in natural waters (with comparisons to other equations)," J. Acoust. Soc. Am. **56**, 1084–1091 (1974).
[48]R. A. L. Antonov J I, T. P. Boyer, A. V. Mishonov, and H. E. Garcia, *World Ocean Atlas 2005, Volume 2: Salinity*, (U.S. Government Printing Office, Washington, DC, 2006).
[49]A. V. M. Locarnini R A, J. I. Antonov, T. P. Boyer, and H. E. Garcia, *World Ocean Atlas 2005, Volume 1: Temperature* (U.S. Government Printing Office, Washington, DC, 2006).
[50]S. Watwood, P. Miller, M. Johnson, P. Madsen, and P. Tyack, "Deep-diving foraging behaviour of sperm whales (*Physeter macrocephalus*)," J. Anim. Ecol. **75**, 814–825 (2006).

Comparison of the properties of tonpilz transducers fabricated with ⟨001⟩ fiber-textured lead magnesium niobate-lead titanate ceramic and single crystals

Kristen H. Brosnan[a] and Gary L. Messing
Department of Materials Science and Engineering, The Pennsylvania State University, 121 Steidle Building, University Park, Pennsylvania 16802

Douglas C. Markley and Richard J. Meyer, Jr.
The Applied Research Laboratory, P.O. Box 30, State College, Pennsylvania 16804

(Received 23 March 2009; revised 4 August 2009; accepted 23 August 2009)

Tonpilz transducers are fabricated from ⟨001⟩ fiber-textured $0.72Pb(Mg_{1/3}Nb_{2/3})O_3-0.28PbTiO_3$ (PMN-28PT) ceramics, obtained by the templated grain growth process, and PMN-28PT ceramic and Bridgman grown single crystals of the same composition. In-water characterization of single element transducers shows higher source levels, higher in-water coupling, and more usable bandwidth for the 81 vol % textured PMN-28PT device than for the ceramic PMN-28PT element. The 81 vol % textured PMN-28PT tonpilz element measured under large signals shows linearity in sound pressure levels up to 0.23 MV/m drive field but undergoes a phase transition due to a lowered transition temperature from the $SrTiO_3$ template particles. Although the textured ceramic performs well in this application, it could be further improved with compositional tailoring to raise the transition temperature and better processing to improve the texture quality. With these improvements textured piezoelectric ceramics will be viable options for medical ultrasound, actuators, and sonar applications because of their ease of processing, compositional homogeneity, and potentially lower cost than single crystal. © *2009 Acoustical Society of America.*
[DOI: 10.1121/1.3238158]

PACS number(s): 43.30.Yj, 43.38.Fx [AJZ] Pages: 2257–2265

I. INTRODUCTION

The tonpilz transducer design is an optimization of the Langevin transducer.[1] This design typically incorporates a heavy tail mass (steel or tungsten), a piezoelectric ceramic stack driven in "33" mode, and a light, flared head mass (magnesium or aluminum). The entire transducer is held under compression by a center bolt; so that during ac drive the compressive stress in the stack is greater than the peak alternating stress, preventing tensile stresses in the piezoelectric stack section during ac drive. During ac drive, maximum displacement occurs at the head mass and the head radiates into the water. The relative masses of the head and tail and the stiffness of the ceramic stack determine the resonance frequency of the transducer.[2] These transducers are used in arrays for active sonar in underwater vehicles.

Currently, tonpilz transducers for underwater vehicles mainly use lead zirconate titanate (PZT) ceramics in the motor section of the transducer. Recently, research on the tonpilz transducers with ⟨001⟩ oriented lead magnesium niobate-lead titanate (PMN-28PT) single crystal show higher source levels and wider bandwidths (BWs) than PZT ceramics due to the higher piezoelectric coefficient and higher electromechanical coupling of the single crystal.[3–5] Textured PMN-32.5PT obtained by the templated grain growth (TGG) process using strontium titanate templates[6] and barium titanate templates[7,8] display better electromechanical properties than the ceramic and a high percentage of the piezoelectric properties of Bridgman grown PMN-28PT single crystals. However, these textured materials have yet to be demonstrated in a transducer design, particularly under large signal.

In this work, oriented PMN-28PT ceramics are fiber-textured in the ⟨001⟩ by the TGG process using a low concentration (5 vol %) of oriented $SrTiO_3$ template crystals.[9,10] The textured ceramics are analyzed by x-ray diffraction, scanning electron microscopy, and electron backscatter diffraction, and incorporated into a tonpilz transducer design. The in-water transducer characteristics are compared for ceramic, 81 vol % textured ceramic, and single crystal PMN-28PT versions of the element.

II. EXPERIMENTAL PROCEDURES

⟨001⟩ textured PMN-28PT blocks of $12\times12\times2$ cm^3 are fabricated by the TGG process.[9,10] Texture is verified using x-ray diffraction. Ceramic PMN-28PT is also fabricated for comparison. For both ceramic and textured PMN-28PT, rings are cut from the ceramic block by sonic milling. The final dimensions of the rings are ~10 mm outer diameter, ~3 mm inner diameter, and ~2 mm thickness. Single crystal PMN-28PT rings of similar dimensions (PMN-PT-28, Morgan Electro Ceramics, Bedford, OH) are obtained commercially.

[a]Author to whom correspondence should be addressed. Electronic mail: brosnan@ge.com. Present address: General Electric Global Research, One Research Circle, Niskayuna, NY 12309.

The ceramic and 81 vol % textured PMN-28PT rings are polished to 15 μm roughness and electroded by screen printing silver ink (6160 Silver Conductor, DuPont Microcircuit Materials, Research Triangle Park, NC) and subsequent firing (850 °C, 10 min). The electroded rings are poled at 1.5 MV/m at 20 °C for 5 min in polydimethylsiloxane (Dow Corning 200® Fluid, Dow Corning, Midland, MI). The poling direction is parallel to the ⟨001⟩ texture axis in the textured rings (through the thickness).

The dielectric constant (ε_{33}) and dielectric loss (tan δ) of the ceramic and textured PMN-28PT rings are measured at 1 kHz with a multifrequency impedance meter (HP4284A Precision *LCR* meter, Agilent Technologies, Inc., Santa Clara, CA) after aging 10 days. The effective coupling (k_{eff}) of the rings is measured using an impedance/gain phase analyzer (HP4194A, Agilent Technologies, Inc., Santa Clara, CA) in conjunction with software to obtain 1 Hz resolution of the series frequency (f_s) and parallel resonance frequency (f_p) of the rings (ZELEMENT V.2.0, Applied Research Laboratory, University Park, PA). Rings with similar k_{eff} and ε_{33} are selected for integration into a tonpilz transducer design.

To aid in the transducer fabrication, models of the tonpilz transducer are created using a graphical user interface (GID V. 7.4.9B, Magsoft Corporation, Ballston Spa, NY) and calculated using ATILA finite element modeling software (ATILA V. 5.2.4, ISEN, Lille, France). The piezoelectric coefficients, compliances, and dielectric properties of the models are assigned in the program to ceramic PMN-28PT, 81 vol % textured PMN-28PT, and single crystal PMN-28PT. These low-field properties are obtained by IEEE standard resonance measurements on shear, longitudinal, disk, and transverse mode samples of ceramic, textured, and single crystal PMN-28PT.[11] The piezoelectric coefficients and elastic coefficients are calculated from the impedance spectra around resonance (f_r) and anti-resonance (f_a). The series resonance and parallel resonance are determined from the maximum of the real part of the conductance (G) and the maximum of the real part of the resistance (R), respectively, in the impedance/frequency scans. The piezoelectric stack length is adjusted in the model so that the transducer resonance frequency matched closely with the desired resonance frequency of the fabricated tonpilz element. The fabricated tonpilz element then incorporates the optimal piezoelectric stack length determined from the ATILA models.

Identical magnesium head masses, tungsten tail masses, brass shims, and epoxies are used for all of the transducers in this study. All parts are cleaned with toluene, ethyl alcohol, and acetone prior to fabrication. Crimped brass electrode shims (thickness=0.10 mm) with lead wires are bonded between the piezoelectric rings with epoxy (Armstrong A-2, Resin Technology Group, Easton, MA). The stacked rings are put under a prestress of 6.9 MPa by tightening the build bolt and monitoring the voltage output from the stack with an electrometer (Keithley 614, Keithley Instruments, Inc., Cleveland, OH). The desired voltage output is estimated from

$$V = \frac{g_{33}\sigma_{33}\, tn\, C}{nC + C_{\text{ref}}}, \tag{1}$$

where g_{33} is the piezoelectric voltage coefficient of the rings in V m/N, σ_{33} is the desired preload stress, t is the ring thickness, n is the number of rings, C is the ring capacitance in nF, and C_{ref} is a parallel reference capacitor in nF (fixed at 17 μF).

The in-air complex electrical impedance of the stack is monitored after each step in the fabrication using an impedance/gain phase analyzer (HP4194A, Agilent Technologies, Inc., Santa Clara, CA). The stack capacitance and dielectric loss are monitored with an impedance analyzer at 1 kHz (HP4284A Precision *LCR* meter, Agilent Technologies, Inc., Santa Clara, CA). To study the effect of preload on the transducer coupling experiments, the bolt is tightened to incrementally increase the preload stress from 0 to 20.7 MPa. After adjusting the preload on the element, the element is aged for 1 day, the capacitance and dielectric loss are measured, the complex electrical impedance frequency sweeps are collected, the bolt is removed, and finally the element is poled (at 1.5 MV/m for 5 min, room temperature) and the process repeated to a maximum of 20 MPa preload stress. The element is poled before each preload stress increment to align the polarization vectors so that each adjustment in preload stress begins with a material that has similar polarization state.

For the in-water tests on the single elements, the tonpilz elements are preloaded to either 6.9 (small signal tests) or 17.2 MPa (high drive tests). The voltage output is monitored on the electrometer while increasing a preload stress of the element (without the bolt) using a home-built press and calibrated strain gauge. The final bolt is tightened to the same voltage output after the load is removed.

The element is then mounted into a cylindrical housing. A *K*-type thermocouple is attached to the outer surface of the piezoelectric stack on the element using the same rapid cure adhesive. The element is suspended in the center of a stainless steel housing end cap. Finally, the element head mass and end cap surfaces are bonded to a neoprene window (of the same area as the end cap) with an instant adhesive (Loctite 410, Henkel Corporation, Rocky Hill, CT).

Small signals in water tests on the single elements are performed in the anechoic water tank at the Applied Research Laboratory at the Pennsylvania State University. The water tank holds 233 kl of water and measures 5.47 m depth × 5.32 width × 7.90 m length. The transducer (in the housing cylinder) and the hydrophone (used to measure the transducer acoustic output) are placed at a depth of 2.43 m. The calibrated hydrophone and the test transducer are separated in the water by 3.16 m. A sinusoidal pulse of 2.0 ms duration is used. A vector signal analyzer is used to maximize the signal to noise ratio and provide precision time gated signal processing (HP89410A, Agilent Technologies, Inc., Santa Clara, CA). A drive level of 30 dB V_{rms} is used for all of the low signal in-water tests. A dc bias is applied to the transducers in the levels of 300 and 600 V in some experiments. Beam patterns are collected at three frequencies around the resonance frequency of the transducer by horizon-

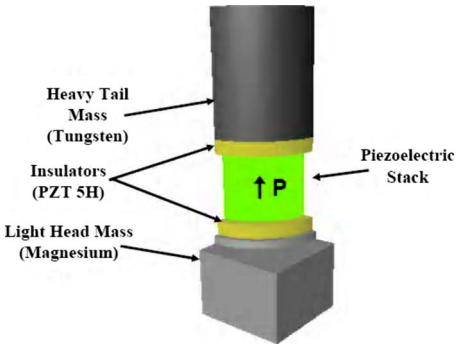


FIG. 1. (Color online) The tonpilz model created using GID graphical interface. Only $\frac{1}{4}$ of the tonpilz is modeled due to symmetry of the element. The polarization direction of the stack is shown by the arrow.

tally rotating the transducer at 360° in the water, and measurements are taken at approximately every degree. The data collected (beam patterns and frequency sweeps) are compensated for the inductance, capacitance, and resistance (*LCR*) of the test cable.

High drive in-water tests on the 81 vol % textured PMN-28PT single element are performed in the high pressure water tank at the Applied Research Laboratory at the Pennsylvania State University. A hydrostatic water pressure of 1.03 MPa is maintained during high drive tests to prevent cavitation. The water tank holds 7.57 kl of water, measures 4.19 m depth × 1.52 m diameter, and the water is maintained at 24 °C. The transducer (in the housing, attached to the lid of the high pressure tank) and the hydrophone (located at the bottom of the tank) are separated by 1.36 m. A sinusoidal pulse of 2.0 ms (0.6% duty cycle) duration is used, generated by a digital tone burst timing generator (Dranetz 658, Dranetz-BMI, Edison, NJ) and amplified by a power amplifier (L-10, Instruments, Inc., San Diego, CA). The drive level is incrementally increased from 30 to 55 dB V_{rms} (0.02–0.37 MV/m) for the high drive tests. A dc voltage bias of $\sqrt{2}$ V_{rms} is applied to the transducers in all high drive experiments. The data collected are compensated for the *LCR* of the test cable. In addition, the frequency sweep data from the small signal anechoic water tank measurements are used to calibrate the high pressure tank data for sound reflections from the walls of the vessel.

III. RESULTS AND DISCUSSION

A. Transducer modeling and fabrication

A model created using the GID graphical interface is shown in Fig. 1. Only $\frac{1}{4}$ of the element is modeled due to the fourfold symmetry of the transducer. This allows for faster computation of the model with the ATILA FEM (finite element model) software. In the models, the center bolt is removed. The model does not account for glue joints, electrodes, or losses (dielectric, mechanical, and piezoelectric). The model is dependent on the accuracy of the full property data sets for each of the materials in the transducer (head mass, tail mass, bolt, insulators, and most importantly, the piezoelectric stack). The property sets used for the piezoelectric stack sections obtained by IEEE standard resonance techniques on different geometry samples are listed in Table I.

TABLE I. Dielectric, piezoelectric, and electromechanical coupling, and compliance coefficients for ceramic, 81 vol % textured ceramic, and single crystal PMN-28PT materials measured by the IEEE resonance technique (cut and poled in ⟨001⟩). These properties are assigned to the piezoelectric stack sections of the ATILA FEA models.

Value	Ceramic PMN-28PT	81 vol % textured PMN-28PT	Single crystal PMN-28PT
Symmetry	∞ m	∞ m	4 mm
ε_{11}^{T}	2230	4200	1550
ε_{33}^{T}	2150	3800	5400
d_{31} (pC/N)	−130	−450	−700
d_{33}	295	940	1540
d_{15}	540	840	165
s_{11}^{E} ($\times 10^{-12}$ m^2/N)	11.4	26.0	52
s_{12}^{E}	−3.8	−9.04	−24.6
s_{13}^{E}	−4.7	−17.1	−26.4
s_{33}^{E}	12.0	33.0	59.9
s_{44}^{E}	38.3	48.0	16.0
s_{66}^{E}	30.4	71.5	28.3
ρ (kg/m^3)	8010	7800	8030

The inputs for the model for the piezoelectric stack are density (ρ), the piezoelectric coefficients (d_{ij}), the dielectric permittivity (ε_{ij}), and the elastic compliance coefficients (s_{ij}). The model can be used to predict vibration modes, the source level (sound pressure level in water at 1 m), and the complex electrical impedance as a function of frequency of the transducer.

The rings used in fabrication of the ceramic and 81 vol % textured PMN-28PT tonpilz elements are fully characterized prior to incorporation into the devices. The k_{eff} of the rings is calculated from

$$k_{\text{eff}} = \sqrt{1 - \left(\frac{f_s^2}{f_p^2}\right)}, \tag{2}$$

where f_s is the frequency of the maximum of the real part of the conductance and f_p is the frequency of the maximum of the real part of the resistance in the complex electrical impedance/frequency scan. The mechanical quality factor (Q_M) is calculated from the radial mode resonance from

$$Q_M = \frac{f_s}{f_1 - f_2}, \tag{3}$$

where f_1 is the frequency of the minimum of the susceptance (B) (also the imaginary part of the admittance) and f_2 is the frequency of the maximum of the susceptance (B). The radial mode is used for this characterization because a clean resonance peak is generated in this mode due to the geometry of the rings. The dielectric data are collected at room temperature and at 1 kHz. All rings are aged over 10 days prior to measurement. The piezoelectric coefficient d_{33} is measured with a Berlincourt meter on all rings after aging, and is 1035 ± 7 and 315 ± 5 pC/N for 81 vol % textured and ceramic PMN-28PT, respectively (Table II). The d_{33} values are comparable to the measured properties of 81 vol % textured PMN-28PT and ceramic PMN-28PT from IEEE resonance method (d_{33}=940 and d_{33}=295 pC/N for 81 vol % textured and ceramic PMN-28PT, respectively, from Table I), al-

TABLE II. Electromechanical and dielectric properties of ceramic, 81 vol % textured ceramic, and single crystal PMN-28PT rings for the piezoelectric stack section of the fabricated tonpilz transducers.

Material	d_{33} (pC/N)	k_{eff}	Q_M	ε_{33}	tan δ
81 vol % textured PMN-28PT ($N=16$)	1035 ± 7	0.63 ± 0.005	145 ± 11	3685 ± 57	0.004 ± 0.0002
Ceramic PMN-28PT ($N=19$)	315 ± 5	0.38 ± 0.005	113 ± 4	1850 ± 48	0.022 ± 0.0007
Single crystal PMN-28PT	1370 ± 75	0.58 ± 0.08	220 ± 45	5500 ± 260	0.003 ± 0.0002

though the ring geometry is not ideal for accurate Berlincourt d_{33} measurement. Table II summarizes the average properties of the piezoelectric rings fabricated (ceramic and 81 vol % textured ceramic PMN-28PT) and bought commercially (single crystal PMN-28PT). In general, the standard deviation of the dielectric constants and electromechanical properties of the rings fabricated is quite small.

Tonpilz elements fabricated from ceramic and textured PMN-28PT rings are shown in Fig. 2. One tonpilz element with eight ceramic PMN-28PT rings is fabricated and one tonpilz element with six textured PMN-28PT rings is fabricated. The thermocouple attached to the outer surface of the rings is shown on the six-ring textured PMN-28PT tonpilz element. The small signal data from these elements are compared to single crystal PMN-28PT tonpilz element data collected at the Applied Research Laboratory in Sec. III C.

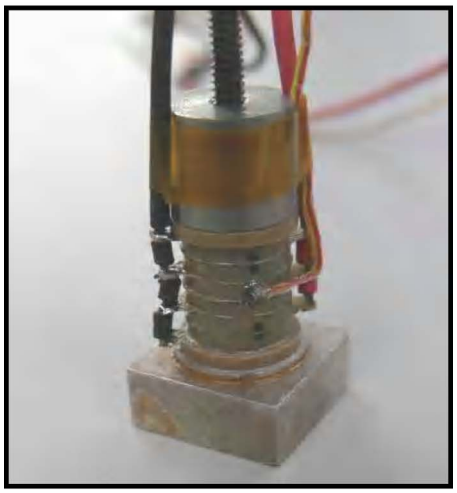
Comparison of the model electromechanical coupling and resonance frequency taken from electrical impedance data to the actual measured in-water data is shown in Table III. The resonance frequencies shown are relative to the measured resonance frequency of the single crystal PMN-28PT tonpilz element. The modeled textured ceramic and measured resonance frequencies are equivalent, which indicates that the materials property data used in the model for the textured PMN-28PT ceramic are accurate. The ceramic PMN-28PT and single crystal PMN-28PT data sets may not be as accurate due to the discrepancies in the modeled and measured resonance frequency values. In general, the tonpilz elements display resonance frequencies near the resonance frequency of the single crystal element. The error in the models is greater than the measured values due to the low frequency resolution in the models.

The models predict a significantly higher in-water coupling for the textured PMN-28PT element ($k_{\text{eff}}=0.74$) compared to the ceramic element ($k_{\text{eff}}=0.55$). However, the electromechanical coupling is lower for the measured transducers than for the model transducers for all cases. The model does not accurately predict the electromechanical coupling because the model does not have a complete set of mechanical, piezoelectric, and dielectric losses for the piezoelectric stacks. The discrepancy in electromechanical coupling could also be a result of the prestress that is applied in the measured transducers that influences the electromechanical properties. The effect of uniaxial stress on the electromechanical coupling in the textured PMN-28PT element is addressed in Sec. III B.

B. Effect of preload stress on 81 vol % textured PMN-28PT transducer in-air properties

In tonpilz designs, the center bolt applies a compressive force on the element. The stress is greater than the peak alternating stress, preventing tensile stresses in the piezoelectric stack section during ac drive.[2] For the tonpilz design, the

(a)



(b)

FIG. 2. (Color online) Tonpilz elements fabricated from (a) eight rings of ceramic PMN-28PT and (b) six rings of 81 vol % textured PMN-28PT (shown with attached *K*-type thermocouple). For both materials, the rings are approximately 10 mm in diameter and 2 mm in thickness.

TABLE III. Comparison of tonpilz in-water electromechanical coupling and resonance frequency [relative to the single crystal PMN-28PT measured resonance frequency ($f_{s\text{-sc}}$)] of modeled and measured transducers.

Element	$f_s/f_{s\text{-sc}}$ (measured)	$f_s/f_{s\text{-sc}}$ (ATILA model)	k_{eff} (measured)	k_{eff} (ATILA model)
Single crystal PMN-28PT	1.00	0.87	0.72	0.80
Textured PMN-28PT	0.73	0.73	0.57	0.74
Ceramic PMN-28PT	0.79	0.97	0.36	0.55

typical preload stress on the single crystal PMN-28PT elements is 17.2 MPa. Textured ceramics used in transducer applications must be operated under a uniaxial compressive stress as well. It is important that the textured ceramics retain high coupling and source levels under a moderate uniaxial compressive stress. In a previous study on textured PMN-PT, Sabolsky[12] suggested that there is ferroelastic switching in the textured ceramics resulting in depolarization with applied stress. Recent studies on stress loading of single crystal PMN-32PT suggest that with no applied bias field, a rhombohedral to orthorhombic phase transition takes place in the range 10–20 MPa, resulting in depolarization of the crystal (in the direction of the stress). With ⟨001⟩ compressive stress, rhombohedral domains are depolarized into the orthorhombic state, which has lower strain in the ⟨001⟩ and spontaneous polarization perpendicular to the [001].[13]

The in-air coupling of the textured PMN-28PT element and the dielectric constant of the stack is shown in Fig. 3 as a function of preload stress. The effect of uniaxial stress on the in-air coupling of the textured PMN-28PT transducer and dielectric properties of the stack is determined by measuring the electrical impedance spectra of the element at 1 V_{rms} after a preload is applied at preload stress levels of $\sigma = 0-20.7$ MPa. Initial application of the preload stress increases the resonance and anti-resonance frequencies of the

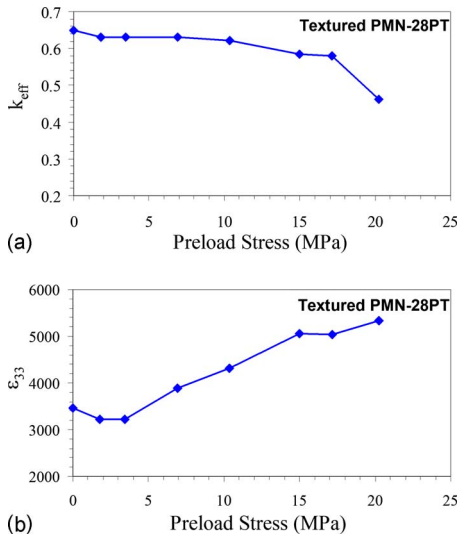


FIG. 3. (Color online) (a) In-air element electromechanical coupling (k_{33}) of the element and (b) dielectric constant (ε_{33}) of the stack as a function of preload stress on the 81 vol % textured PMN-28PT tonpilz element.

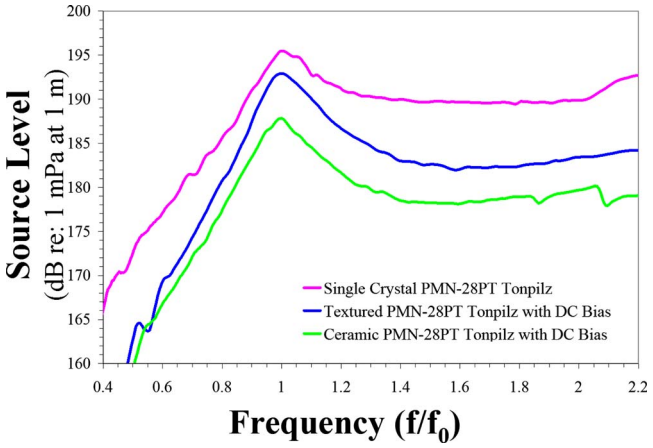


FIG. 4. (Color online) (a) Source level in water for the three transducer cases normalized at 0.10 MV/m measured from small signal (30 dB V_{rms}) measurements in the anechoic tank, and frequencies have been normalized to the resonance frequency of each transducer.

transducer. Above 6.9 MPa, only the resonance frequency increases with higher preload stress, but the anti-resonance frequency remains unchanged.

In Fig. 3(a), at $\sigma = 0$ MPa preload stress, the element coupling is high, with a $k_{\text{eff}} = 0.66$. The coupling remains high at $k_{\text{eff}} = 0.63$ at $\sigma = 6.9$ MPa. However, the in-air coupling decreases to $k_{\text{eff}} = 0.58$ at $\sigma = 17.1$ MPa, and to $k_{\text{eff}} = 0.46$ at $\sigma = 20.7$ MPa. The element retained high coupling up to the prestress level desired for this tonpilz design ($\sigma = 17.2$ MPa). It should be noted, however, that there is little room for error in applying the prestress without an orthorhombic phase transition. Overshooting a preload of 17.2 MPa would most likely result in poor element coupling from depolarization of the textured PMN-28PT stack. In comparison, ⟨001⟩ single crystal PMN-30PT and ceramic PMN-30PT materials have shown high coupling (k_{33}) greater than 0.91 and 0.60 under uniaxial stresses up to 40 MPa and above 60 MPa, respectively.[14,15] Considering the 10–20 MPa limit on single crystal PMN-32PT, the effect of prestress on the effective coupling appears to be a function of composition.

The dielectric constant (calculated from capacitance measurements at 1 kHz) of the stack increases with increasing uniaxial stress, from $\varepsilon_{33} = 3410$ at 0 MPa to $\varepsilon_{33} = 5325$ at $\sigma = 20.7$ MPa. Previous studies in fine grain ceramic PMN-30PT (average grain size = 1.72 μm) show a slight decrease in the dielectric constant with uniaxial stress (up to 230 MPa).[16,17] These trends can be explained by a rhombohedral to orthorhombic phase transition, which allows for rotation of the polarization vectors under an applied stress.[13,18] This effect occurs at lower compressive stress than single crystal PMN-28PT due to the presence of Sr^{2+} in the textured ceramic (which is not present in the single crystal); Sr^{2+} substitutes for Pb^{2+} in PZT and PMN ceramics. Similar observations of domain switching under applied uniaxial stress are observed in soft PZT ceramics.[19] It should be noted, however, that the above measurements on the textured ceramic element are conducted at low-field $E_{\text{ac,PP}} = 8.0 \times 10^{-5}$ MV/m, 0 MV/m dc bias, which is different from most data in literature in which an $E_{\text{ac,PP}} = 1.5$ MV/m and dc bias of 0.70 MV/m, which results in higher uniaxial stresses

TABLE IV. Electromechanical coupling (k_{eff}), mechanical quality factor (Q_M), and BW for the three transducer cases at 30 dB V_{rms} drive under 0, 0.20, and 0.40 MV/m dc bias fields. BW comparison for all transducer cases is calculated at arbitrary intercept of $VA/P_{acs}=7$ V A/W.

Element	0 MV/m dc bias		0.20 MV/m dc bias		0.40 MV/m dc bias		BW (%)
	k_{eff}	Q_M	k_{eff}	Q_M	k_{eff}	Q_M	
Single crystal PMN-28PT	0.72	5.04	⋯	⋯	⋯	⋯	100
Textured PMN-28PT	0.57	5.53	0.61	5.56	0.64	5.73	66.5
Ceramic PMN-28PT	0.36	6.64	0.39	6.64	0.44	6.6	31.4

needed to induce domain switching.[14,15,18–20] Nevertheless, the results suggest that there are limitations on the compressive preload stress (and thus the drive levels) for the 81 vol % textured ceramic. This may be overcome in the future through further compositional tailoring.

C. Small signal in-water transducer characterization

The tonpilz elements are mounted into a housing cylinder for in-water testing in the anechoic tank at the Applied Research Laboratory. In-water small signal tests are conducted on three transducers: single crystal PMN-28PT, 81 vol % textured PMN-28PT, and ceramic PMN-28PT. The relative source level in water for the three transducer cases is shown in Fig. 4. The frequency for each case is normalized for the resonance frequency of each transducer (f/f_0), and the source level is normalized for 0.10 MV/m in Fig. 4. From this figure, the single crystal PMN-28PT element displays the highest source level over the widest frequency range, and the textured ceramic is much better than the ceramic element.

Table IV shows the electromechanical coupling, mechanical quality factor, and BW for the same three transducer cases. With an application of a dc bias, the electromechanical coupling increases and mechanical quality factor decreases for the textured and ceramic PMN-28PT elements. The textured PMN-28PT element reaches a $k_{eff}=0.64$ with 0.40 MV/m dc bias in the low signal in-water tests. This result is promising as the decrease in electromechanical coupling from the uniaxial prestress is negated by a modest dc bias field. The dc bias may stabilize the rhombohedral phase of the ceramic and the 81 vol % textured PMN-28PT, increasing the electromechanical coupling of respective tonpilz elements.[13]

The beam patterns of the ceramic and textured PMN-28PT elements are shown in Fig. 5. The directivity index (DI), or the measure of the acoustic beam relative to an

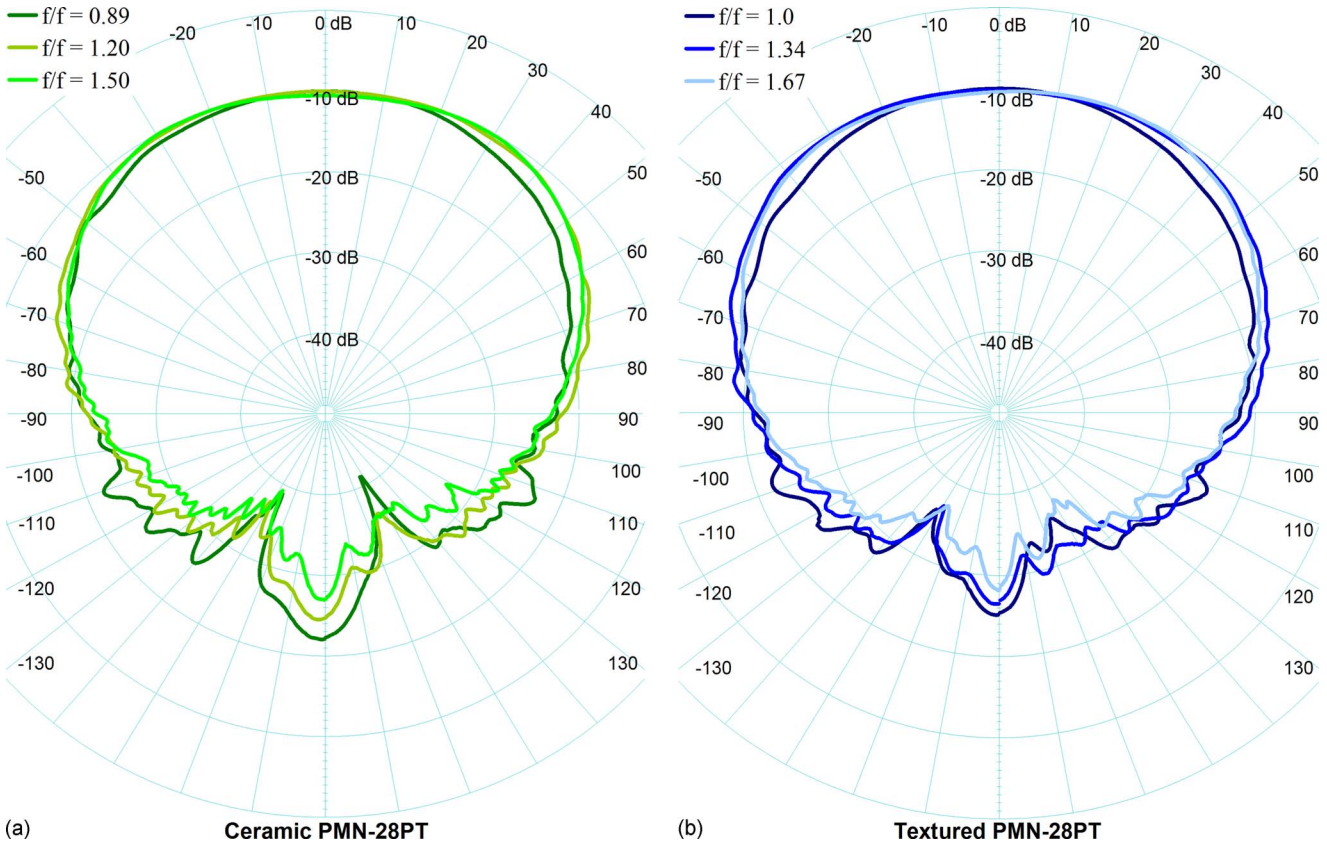


FIG. 5. (Color online) Measured beam patterns showing transducer directivity for (a) ceramic PMN-28PT and (b) textured PMN-28PT transducers. Data in both plots normalized to −10 dB.

omni-directional source, is estimated from the beam patterns.[2] The beam patterns show the relative acoustic output (normalized to −10 dB) as the transducer is rotated horizontally at 360°. Beam patterns are measured at several frequencies, as shown in Fig. 5. In order to estimate the acoustic output power and efficiency of the transducer over the entire frequency range, the DI at each frequency is needed. The following equations for a piston source are used to estimate the DI as a function of frequency:[21]

$$\mathrm{DI} = 10 \log\left[\frac{(kr)^2}{1 - \frac{J_1(2kr)}{kr}}\right], \tag{4}$$

$$\lambda = \frac{v_{sw}}{f} \quad \text{and} \quad k = \frac{2\pi}{\lambda}, \tag{5}$$

where f is the frequency, v_{sw} is the speed of sound in water (1500 m/s), λ is the wavelength, r is the effective radius of the head mass, and $J_1(2kr)$ is the Bessel function of the first kind and first order. At low frequencies, this approximation is less accurate since the wavelength (λ) is greater than the effective radius of the head mass (r).[2] The calculated DI ranges from 4.4 to 6.6 dB for $f/f_0 = 1.00-1.67$ (relative to the resonance frequency of the textured PMN-28PT element), respectively, using the above approximation for a piston source. At higher frequencies, $f/f_0 \geq 1.34$, the calculated and measured DI are in good agreement using the above approximation.

The acoustic power (P_{acs}) is calculated from the source level and DI as a function of frequency from[2]

$$P_{acs} = 10^{(\mathrm{SL}-\mathrm{DI}-170.9/10)}, \tag{6}$$

where SL is the sound pressure level of the element at 0.10 MV/m (relative to 1 μPa at 1 m), P_{acs} is the acoustic power output at 0.10 MV/m, and DI is the directivity index. Both the SL and DI are frequency dependent, and thus the acoustic power output is also frequency dependent. All three elements are fabricated with the same head mass; thus the DI is taken to be equal for each transducer case.

A good figure of merit for transducers is the apparent electrical power input (VA) divided by the acoustic power output (P_{acs}), the so called "transmit system performance" or TSP. A $VA/P_{acs} = 1$ represents 100% efficiency.[2] TSP is used instead of efficiency [which is the ratio of the acoustic power output over the (real) electrical power input] because amplifiers have to deliver both real and reactive power; thus, the apparent electrical power input in the TSP includes both the real and reactive power. The TSP of the three transducer cases is shown in Fig. 6. Based on an arbitrary VA/P_{acs} ratio of 7, the plot shows that the single crystal has the broadest operating BW. The BW (relative to resonance frequency of each transducer) is shown in Table IV. The single crystal element showed the highest BW (100%) and the ceramic element the lowest (31.4%). The textured tonpilz elements showed twice the usable BW (i.e., 66.5%) relative to the ceramic PMN-28PT tonpilz element. A more accurate comparison of these materials in the tonpilz design will be made at the array level.

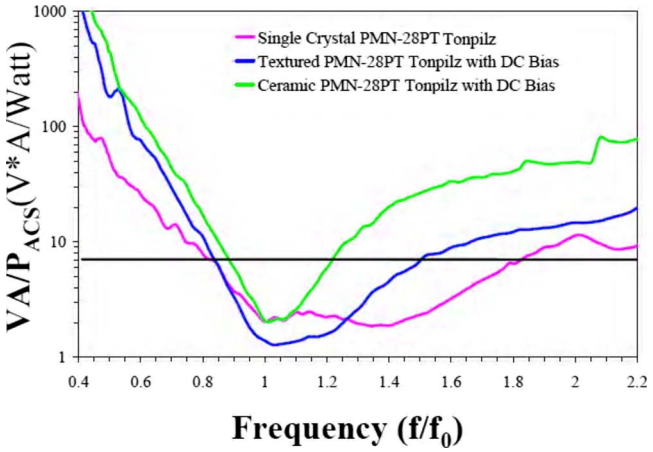


FIG. 6. (Color online) Efficiency as a function of frequency for the three transducer cases. Intercept line drawn at arbitrary value of $VA/P_{acs} = 7$ V A/W shows the relative difference in operational BW of the three transducer cases.

D. High drive transducer characterization

In Sec. III C, the results are scaled linearly based on low-field measurements. High drive characterization aids in modeling and in the prediction of device performance under typical operating conditions. In this section, linearity is tested by increasing the drive level and monitoring the transducer output. The 81 vol % textured PMN-28PT element is tested under high drive conditions in the high pressure tank at the Applied Research Laboratory. In the high drive test, the duty cycle is kept low (0.6%) and the complex electrical impedance/frequency data are collected at drive levels from $E_{ac,PP} = 0.02$ to 0.37 MV/m with dc bias (E_{dc}) of $E_{dc} = \sqrt{2}E_{ac}$. The complex electrical impedance for these conditions is collected and the impedance magnitude shown in Fig. 7(a). The shift in resonance frequency to lower frequencies is more clearly shown in Fig. 7(b), with a Δf_s of −19% at 0.37 MV/m drive. This shift is evidence of the "soft" behavior of the 81 vol % textured PMN-28PT. Single crystal PMN-28PT changes −10% over the same conditions.

During measurements at all drive levels, the increase in the temperature of the element is negligible. At $E_{ac,PP} = 0.37$ MV/m, the experiment is halted due to current spikes during data collection. Inspection of the element after testing showed a carbon trace on the outside of one of the rings, indicating short circuit conditions from dielectric breakdown. The element itself is found to be intact, and low signal in-air complex electrical impedance data are identical to the electrical impedance/phase angle data prior to the high drive tests. Future tests should consider a conformal coating to help prevent arcing.

The electromechanical coupling of the textured PMN-28PT element increases with drive level and plateaus at $k_{eff} = 0.69$ [Fig. 8(a)]. The mechanical quality factor decreases slightly to $Q_M = 4.35$ with increase in drive level. The maximum source level increases with drive level and follows a linear relationship with drive (in dB) for electric fields up to 0.23 MV/m [Fig. 8(b)]. The source level obtained at 0.10

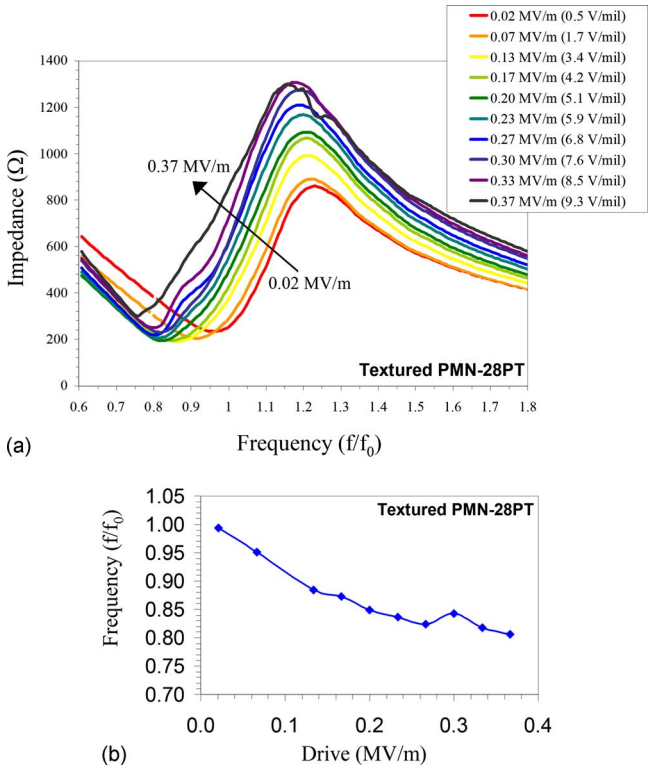


FIG. 7. (Color online) (a) Electrical impedance magnitude and (b) shift in resonance frequency for textured PMN-28PT tonpilz element during high drive conditions ($E_{ac,PP}$=0.02–0.37 MV/m and $E_{dc}=\sqrt{2}E_{ac}$).

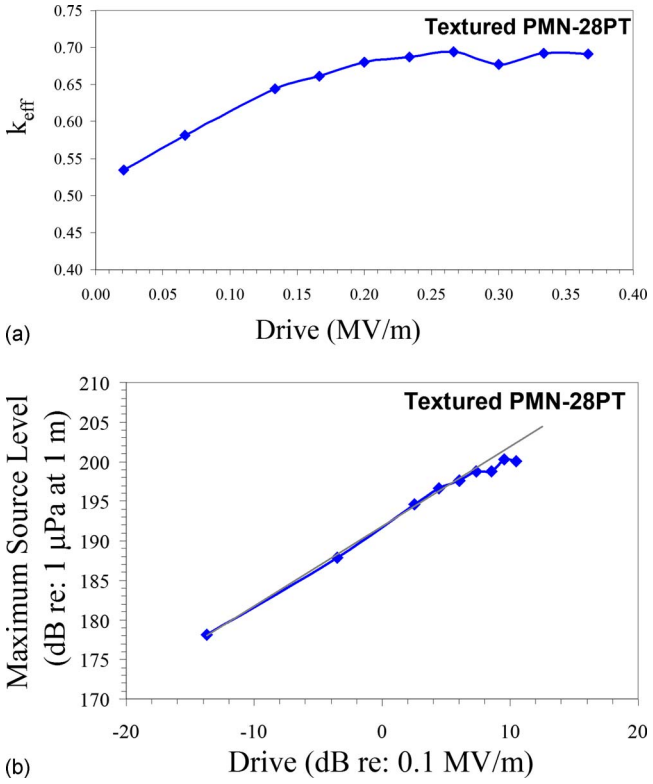


FIG. 8. (Color online) (a) Increase in electromechanical coupling of the textured PMN-28PT tonpilz element with increase in drive level. (b) Maximum source level as a function of drive level in dB (relative to 0.10 MV/m) showing linearity (gray line) in source level with drive for textured PMN-28PT six-ring tonpilz element.

MV/m (SL=191.9 dB) is very close to the source level estimated by the small signal measurements in water, field normalized to 0.10 MV/m (SL=192.8 dB) from Fig. 4.

IV. SUMMARY

In this work, ⟨001⟩ fiber-textured PMN-28PT ceramics of 81 vol % textured fraction are tested in an existing tonpilz transducer design, and performance is compared to ceramic and single crystal PMN-28PT. The in-water transducer characterization showed the textured ceramic to have higher source levels than ceramic PMN-28PT tonpilz elements. High drive tests on the 81 vol % textured PMN-28PT tonpilz element are performed to observe device performance under normal operating conditions. The textured PMN-28PT element showed linearity in source level as a function of drive field up to 0.23 MV/m. The maximum electromechanical coupling obtained by the 81 vol % textured PMN-28PT transducer under high drive conditions is k_{eff}=0.69.

The stress induced phase transition of 81 vol % textured PMN-28PT limits the output power of textured transducers due to the stresses generated during ac drive. The use of $SrTiO_3$ tabular templates to induce texturing is known to reduce the transition temperature of PMN-PT. Thus, transducers made from these textured ceramics will not be able to overcome the deficit in d_{33} (compared to single crystal PMN-28PT) by using higher drive levels than can be achieved with single crystal PMN-28PT. However, a modest dc bias can be used to prevent the stress induced phase transition from the rhombohedral to orthorhombic phase in 81 vol % textured PMN-28PT tonpilz elements. Clearly, template particles that do not reduce the transition temperature but still induce high texture fraction are needed to fabricate textured PMN-PT ceramics comparable to single crystal.

Data presented here show the viability of using the TGG process to enhance the electromechanical response of piezoelectric ceramics and underwater transducer performance. While single crystal still provides the largest electromechanical response, other applications may benefit greatly with TGG materials; especially considering the significantly lower fabrication costs relative to Bridgman grown single crystals.

ACKNOWLEDGMENTS

This paper is based on work supported by the Office of Naval Research, YIP, Grant No. N00014-03-1-0481 and the Applied Research Laboratory Exploratory and Foundational Research Program.

[1]F. A. Fischer, *Fundamentals of Electroacoustics*, 1st ed. (Interscience, New York, 1955).
[2]D. Stansfield, *Underwater Electroacoustic Transducers* (Peninsula, Los Altos Hills, CA, 1991), pp. 179–195.
[3]J. M. Powers, M. B. Moffett, and F. Nussbaum, "Single crystal naval transducer development," presented at the Proceedings of the 12th IEEE International Symposium on Applications of Ferroelectrics (2000).
[4]K. A. Snook, P. W. Rehrig, W. S. Hackenberger, X. Jiang, R. J. Meyer, and D. Markley, "Advanced piezoelectric single crystal based transducers for naval sonar applications," Proc. SPIE **5761**, 263–271 (2005).
[5]R. J. Meyer, Jr., T. C. Montgomery, and W. J. Hughes, "Tonpilz transducers designed using single crystal piezoelectrics," presented at the Oceans 2002 IEEE/MTS, Biloxi, MS (2002), pp. 29–31.

[6]S. Kwon, E. M. Sabolsky, G. L. Messing, and S. Trolier-McKinstry, "High strain, ⟨001⟩ textured $0.675Pb(Mg_{1/3}Nb_{2/3})O_3-0.325PbTiO_3$ ceramics: Templated grain growth and piezoelectric properties," J. Am. Ceram. Soc. **88**, 312–317 (2005).
[7]E. M. Sabolsky, A. R. James, S. Kwon, S. Trolier-McKinstry, and G. L. Messing, "Piezoelectric properties of ⟨001⟩ textured $Pb(Mg_{1/3}Nb_{2/3})O_3-PbTiO_3$ ceramics," Appl. Phys. Lett. **78**, 2551–2553 (2001).
[8]E. M. Sabolsky, S. Trolier-McKinstry, and G. L. Messing, "Dielectric and piezoelectric properties of ⟨001⟩ fiber-textured $0.675Pb(Mg_{1/3}Nb_{2/3})O_3-0.325PbTiO_3$ ceramics," J. Appl. Phys. **93**, 4072–4080 (2003).
[9]K. H. Brosnan, G. L. Messing, R. J. Meyer, and M. D. Vaudin, "Texture measurements in ⟨001⟩ fiber-oriented PMN-PT," J. Am. Ceram. Soc. **89**, 1965–1971 (2006).
[10]K. H. Brosnan, S. F. Poterala, R. J. Meyer, S. Misture, and G. L. Messing, "Templated grain growth of ⟨001⟩ textured PMN-28PT using $SrTiO_3$ templates," J. Am. Ceram. Soc. **92**, S133–S139 (2009).
[11]B. Jaffe, W. R. Cook, and H. Jaffe, *Piezoelectric Ceramics* (Academic, New York, 1971).
[12]E. M. Sabolsky, "Grain-oriented $Pb(Mg_{1/3}Nb_{2/3})O_3-PbTiO_3$ ceramics prepared by templated grain growth," Ph.D. thesis, The Pennsylvania State University, University Park, PA (2001).
[13]E. A. McLaughlin, T. Q. Liu, and C. S. Lynch, "Relaxor ferroelectric PMN-32%PT crystals under stress, electric field and temperature loading: II-33-mode measurements," Acta Mater. **53**, 4001–4008 (2005).
[14]D. Viehland, L. Ewart, J. Powers, and J. F. Li, "Stress dependence of the electromechanical properties of ⟨001⟩-oriented $Pb(Mg_{1/3}Nb_{2/3})O_3-PbTiO_3$ crystals: Performance advantages and limitations," J. Appl. Phys. **90**, 2479–2483 (2001).
[15]D. Viehland, J. F. Li, K. Gittings, and A. Amin, "Electroacoustic properties of ⟨110⟩-oriented $Pb(Mg_{1/3}Nb_{2/3})O_3-PbTiO_3$ crystals under uniaxial stress," Appl. Phys. Lett. **83**, 132–134 (2003).
[16]R. Yimnirun, "Contributions of domain-related phenomena on dielectric constant of lead-based ferroelectric ceramics under uniaxial compressive pre-stress," Int. J. Mod. Phys. B **20**, 3409–3417 (2006).
[17]R. Yimnirun, M. Unruan, Y. Laosiritaworn, and S. Ananta, "Change of dielectric properties of ceramics in lead magnesium niobate-lead titanate system with compressive stress," J. Phys. D **39**, 3097–3102 (2006).
[18]D. Viehland, "Effect of uniaxial stress upon the electromechanical properties of various piezoelectric ceramics and single crystals," J. Am. Ceram. Soc. **89**, 775–785 (2006).
[19]A. B. Schaufele and K. H. Hardtl, "Ferroelastic properties of lead zirconate titanate ceramics," J. Am. Ceram. Soc. **79**, 2637–2640 (1996).
[20]D. Viehland, J. Powers, L. Ewart, and J. F. Li, "Ferroelastic switching and elastic nonlinearity in ⟨001⟩-oriented $Pb(Mg_{1/3}Nb_{2/3})O_3-PbTiO_3$ and $Pb(Zn_{1/3}Nb_{2/3})O_3-PbTiO_3$ crystals," J. Appl. Phys. **88**, 4907–4909 (2000).
[21]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 4th ed. (Wiley, Hoboken, NJ, 2000).

The dependence of the moving sonoluminescing bubble trajectory on the driving pressure

Rasoul Sadighi-Bonabi,[a] Reza Rezaei-Nasirabad, and Zeinab Galavani
Department of Physics, Sharif University of Technology, Tehran 11365-91, Iran

(Received 21 December 2008; revised 24 June 2009; accepted 29 June 2009)

With a complete accounting of hydrodynamic forces on the translational-radial dynamics of a moving single-bubble sonoluminescence, temporal evolution of the bubble trajectory is investigated. In this paper, by using quasi-adiabatic evolution for the bubble interior, the bubble peak temperature at the bubble collapse is calculated. The peak temperature changes because of the bubble translational motion. The numerical results indicate that the strength of the bubble collapse is affected by its translational movement. At the bubble collapse, translational movement of the bubble is accelerated because of the increase in the added mass force on the bubble. It is shown that the magnitude of the added mass force rises by the increase in the amplitude of the driving pressure. Consequently, the increase in added mass force results in the longer trajectory path and duration.
 [DOI: 10.1121/1.3183413]

PACS number(s): 43.35.Ei, 43.35.Hl [DLM] Pages: 2266–2272

I. INTRODUCTION

Single bubble sonoluminescence (SBSL) is the light emission from a trapped collapsing bubble in a standing ultrasound field in liquid.[1] Trapping of the bubble is caused by the primary Bjerknes force and is a combined effect of the sound field and nonlinear bubble oscillations.[2] The balance between the time-averaged primary Bjerknes force and the time-averaged buoyancy force is the reason for the bubble trapping near the pressure antinode.[3,4]

Various aspects of sonoluminescing (SL) bubbles have been studied including the role of liquid viscosity in the dynamics of a SL bubble.[5,6] It is observed that in high viscous liquids, sonoluminescence can occur in a single moving bubble.[7–14] In this phenomenon, which is so called moving single bubble sonoluminescence (m-SBSL),[7] collective effects of the components of the hydrodynamic force[15] on an oscillating bubble cause a translational movement of a bubble in a trajectory.[8,13,14]

Based on the derived expression for the hydrodynamic force on a bubble of changing size in an incompressible liquid,[15] Reddy and Szeri[16] studied the coupled dynamics of translation and the collapse of acoustically driven microbubbles in water. Following their fundamental work, the equation of the moving bubble trajectory was extracted by Toegel *et al.*[14] They introduced the history force as the origin of the bubble translational motion. Although they succeeded partially to show the agreement of their results with the experimentally reported bubble velocity and domain of trajectory,[7] due to the assumption of isothermal evolution of the gas in the bubble, their calculated phase diagram for a SL bubble in water shows a contradiction to what is observed in the experiments at high driving pressures. Therefore, to investigate the characteristics of a SL bubble, other models should be considered for the bubble interior.

Homogeneous van der Waals gas without heat and mass exchange[17] or more sophisticated models that account for the effects of liquid vapor and the endothermic chemical reactions[18–21] can also be used for the sonoluminescence phenomenon. Depending on the requested parameters regarding the dynamics of sonoluminescence, simpler models can be implemented.

In the present work, we are interested in the investigation of the temporal effects of hydrodynamic forces on the translational-radial dynamics of a bubble under moving sonoluminescence condition. Based on the bubble parameters such as the domain of trajectory, we have found that the quasi-adiabatic evolution[17] at the bubble interior is a suitable model. This model gave us the ability to calculate the effect of the bubble translational motion on its peak temperature. Our calculations show that spatial movements of the SL bubble affects the strength of the bubble contraction and its peak temperature at the instant of the bubble collapse in the acoustic cycles. Also at this instant, rapid increase in the wall velocity induces an enhanced momentum on the bubble movement through the increase in the added mass force on it. We numerically examined the effects of the increase in the amplitude of the driving pressure on the bubble trajectory. The results indicate that the average of the added mass force maxima rises when the amplitude of the driving pressure increases.

II. THE METHOD

To model the radial and the translational motions for the bubble, the influence of the full hydrodynamic forces[15] on the bubble should be taken into account. Magnaudet and Legendre[15] presented two asymptotic expressions for the hydrodynamic force on a moving bubble in a liquid. In attempts to formulate the equation of the SL bubble motion,[16,20,22–24] those asymptotic expressions are combined and the path of a collapsing bubble trajectory in high viscous liquid is calculated by Toegel *et al.*[14] They assumed an isothermal model

[a]Author to whom correspondence should be addressed. Electronic mail: sadighi@sharif.ir

for the interior evolution of the bubble which should be replaced with more sophisticated models for the investigation of bubble dynamics under sonoluminescence condition. The most appropriate model is a hydro-chemical model which we used in comparing the dynamics of a moving SL bubble in sulfuric acid to the dynamics of a SL bubble in pure water.[25] Since the calculation of the heat and the mass transfer through the bubble wall and the mechanism of sonoluminescence light emission is not the aim of this work, we used a simple quasi-adiabatic model for the gas inside the SL bubble.[17]

The selected model of the bubble radial evolution, the bubble translational equation, and the components of the hydrodynamic forces are presented in this paper.

A. The equation of the bubble radial oscillation

We used the well-known Rayleigh–Plesset equation[26] to describe the bubble radial evolution:

$$\left(1-\frac{\dot{R}}{c}\right)R\ddot{R}+\frac{3}{2}\dot{R}^2\left(1-\frac{\dot{R}}{3c}\right)=\left(1+\frac{\dot{R}}{c}\right)\frac{1}{\rho}(p_g-p_{\mathrm{ac}}-p_o)+\frac{R\dot{P}_g}{\rho c}-\frac{4\nu\dot{R}}{R}-\frac{2\sigma}{\rho R}, \tag{1}$$

where R, c, ρ, ν, and σ are the bubble radius, sound velocity in the liquid, liquid density, kinetic viscosity (shear viscosity/density) of the liquid and the surface tension at the bubble wall, respectively. P_o is the ambient pressure in the liquid, and $P_{\mathrm{ac}}=-p_a\sin(\omega t)(1-(\pi^2|\mathbf{x}|^2/6R_{fl}^2))$ is the acoustic pressure field around the bubble[14] with $R_{fl}=3$ cm as the resonator's radius.[24] For SBSL condition, the isotropic oscillatory pressure is assumed around a very small bubble. Neglecting the sound radiated by the bubble itself, the velocity potential far from the bubble is a standing wave and the acoustic mode is excited by the transducer. For a trapped bubble in the central antinode of the sound field, we do not require the entire spatial structure of this mode, but only the field close to the bubble. Since the bubble is much smaller than the sound wavelength, this sound field will be independent of $\mathbf{x}$, the radial distance from the central antinode in the standing sound field in the resonator. To apply the spatial distribution of the pressure on the moving SL bubble dynamics in standing sound field, we consider the spherical symmetry and a node at the flask's surface. The pressure field of the lowest resonance mode, P_{ac}, at a small distance from the center of the flask is given in Eq. (1). However, spatial changes in P_{ac} occur on a length scale of the order of the resonator's radius, the typical change in position of the bubble during 1 cycle turns out to be much smaller than sub-microns (much smaller than the standing sound wavelength) such that the assumption of the spatial homogeneity for driving of the Rayleigh–Plesset equation is accurate.

The radial equation is closed by the equation of pressure, P_g, in the bubble. In the present model, the internal gas pressure is calculated by a van der Waals type process equation:[17,27]

$$\dot{P}_g(R,t)=\frac{d}{dt}P_g[R(t)]=-\gamma(R,\dot{R},T)\frac{3R^2\dot{R}}{R^3-h^3}p_g, \tag{2}$$

Here, $h=R_o/8.86$ is the van der Waals hard-core radius and γ, the polytropic exponent, is a transition function[28] from the isothermal behavior to the adiabatic behavior of the bubble interior. For the gas pressure inside the bubble, P_g, we used the excluded volume van der Waals equation of the state:

$$p_g\frac{4\pi}{3}(R^3-h^3)=\frac{4\pi}{3}R_0^3\nu_m\Re T, \tag{3}$$

where $\Re$ is the ideal gas constant and ν_m is the specific molar volume under normal conditions. Using the above equation for the bubble's interior pressure, the following differential equation for the bubble temperature is obtained:

$$\dot{T}=-[\gamma(R,\dot{R},T)-1]\frac{3R^2\dot{R}}{R^3-h^3}T-\chi_g\frac{T-T_o}{R^2}, \tag{4}$$

where T_o is the liquid ambient temperature and χ_g is the thermal diffusivity of the gas inside the bubble.

To compute $\gamma(R,\dot{R},T)$ in the model, we used the simple but useful equation of Hilgenfeldt *et al.*:[17]

$$\gamma(\mathrm{Pe})=1+(\Gamma-1)\exp\left(-\frac{A}{(\mathrm{Pe})^B}\right), \tag{5}$$

with parameters $A\approx 5.8$ and $B\approx 0.6$. Using the time dependent, instantaneous Peclet number $\mathrm{Pe}(t)$:

$$\mathrm{Pe}=\mathrm{Pe}(t)=R(t)|\dot{R}(t)|/\chi_g(R,T), \tag{6}$$

$\gamma(\mathrm{Pe})$ can be used for the strong collapse of a moving SL bubble. For a very large part of the driving cycle, the bubble follows an isothermal behavior $[\gamma(\mathrm{Pe}\to 0)\to 1]$ due to the small Peclet number. Significant deviation from the isothermal behavior only occurs in the vicinity of the collapse, where Pe increases due to the rapid bubble wall velocity; consequently, the temperature increases dramatically under compression.

χ_g, the thermal diffusion of the gas inside the bubble, is calculated based on the Enskog theory of dense gases.[29]

B. Equation of bubble translational motion

Translational motion of a SL bubble has been calculated from full force-balanced translational-radial dynamics of the m-SBSL:[14]

$$R(t)^3\dot{\mathbf{v}}=\frac{d}{dt}[(18\nu R+3R^2\dot{R})(\mathbf{u}-\mathbf{v})+3R^3\dot{\mathbf{u}}-2R^3\mathbf{g}]-3R^2\dot{R}\dot{\mathbf{v}}+3\nu\frac{\Theta_r\Theta_t}{R^2}[(6\nu R+3R^2\dot{R})(\mathbf{u}-\mathbf{v})+3R^3\dot{\mathbf{u}}-2R^3\mathbf{g}-R^3\dot{\mathbf{v}}], \tag{7}$$

where $\mathbf{u}$ is the velocity field of the standing acoustic wave, $\mathbf{v}$ is SL bubble velocity vector relative to an inertial frame, $\mathbf{g}$ is the gravitational acceleration, and R is the temporal radius of the SL bubble. Equation (7) is obtained from equating the two asymptotic expressions of hydrodynamic force on a moving bubble into zero.[14]

In the case where the radial Reynolds number $\mathrm{Re}_r = R|\dot{R}|/\nu \geq 1$ or translational Reynolds number $\mathrm{Re}_t = R|\mathbf{u}-\mathbf{v}|/\nu \geq 1$, the hydrodynamic force on the bubble is given by

$$\mathbf{F}(t) = 12\pi\rho\nu R(t)\mathbf{U}(t) + \frac{2}{3}\pi\rho\left\{\frac{d[R(t)^3\mathbf{U}(t)]}{dt} + 2R(t)^3\frac{d\mathbf{U}(t)}{dt}\right\}. \quad (8)$$

Whereas if $\mathrm{Re}_r \ll 1$ and $\mathrm{Re}_t \ll 1$, the above mentioned force is

$$\mathbf{F}(t) = 4\pi\rho\nu R(t)\mathbf{U}(t) + \frac{2}{3}\pi\rho\left\{\frac{d[R(t)^3\mathbf{U}(t)]}{dt} + 2R(t)^3\frac{d\mathbf{U}(t)}{dt}\right\} + 8\pi\rho\nu\int_0^t \exp\left[9\nu\int_\tau^t R(t')^{-2}dt'\right] \times \mathrm{erfc}\left[\sqrt{9\nu\int_\tau^t R(t')^{-2}dt'}\right]\frac{d[R(\tau)\mathbf{U}(\tau)]}{d\tau}d\tau. \quad (9)$$

In both equations, $\mathbf{U}(t)=\mathbf{u}(t)-\mathbf{v}(t)$ is the relative velocity of the bubble. The components of the hydrodynamic force are introduced in Sec. II C. Since the crossover between the two expressions occurs at critical Reynolds numbers[15] $\mathrm{Re}_{r,\mathrm{crit}}=7$ and $\mathrm{Re}_{t,\mathrm{crit}}=0.5$, one can use $\Theta_r = 1/(1+(\mathrm{Re}_r(t)/\mathrm{Re}_{r,\mathrm{crit}}(t))^4$ and $\Theta_t = 1/(1+(\mathrm{Re}_t(t)/\mathrm{Re}_{t,\mathrm{crit}}(t))^4$; the switches turn on the history force effect on the bubble translational-radial dynamics for sufficiently small Reynolds numbers.[17] The effect of the history force on the dynamics of the moving SL bubble vanishes when the product of Θ_r and Θ_t equals to zero. This occurs when the bubble radial or translational velocity increases rapidly which leads to a rapid increase in the radial and the translational Reynolds number.

C. Components of the hydrodynamic force on the bubble

The effect of unsteady force on a spherical particle has long been an area of interest.[30] According to Magnaudet and Legendre,[15] the components of the hydrodynamic force on the moving bubble are as follows.

(1) $\mathbf{F}_{\mathrm{bouy}}=4/3\pi R(t)^3\rho\mathbf{g}$, the buoyancy force, which is always directed against the gravitational field.

(2) $\mathbf{F}_{\mathrm{Bj}}=4/3\pi\langle R(t)^3\nabla p(x,t)\rangle$, the primary Bjerknes force, which is the acoustic radiation force. For a trapped bubble, it is parallel to the gravitational acceleration vector and its net horizontal component must vanish, where $\langle\ldots\rangle$ denotes the time averaging over a period of the acoustic field and $p(x,t)$ is the acoustic pressure around the bubble.

(3) $\mathbf{F}_{\mathrm{mass}}=-2/3\pi\rho(d[R(t)^3\mathbf{U}(t)]/dt)$, added mass force, which is exerted by the flow on the volume occupied by the bubble. $\mathbf{U}(t)=\mathbf{u}(t)-\mathbf{v}(t)$ is the relative velocity of the bubble. $\mathbf{u}$, the velocity of the bubble, is obtained from a balance of the fluid momentum in the far field (in the acoustic limit): $\rho\partial_t\mathbf{u}=-\nabla p_{\mathrm{ac}}$, yielding $\mathbf{u}=[(\pi^2 x_i p_{\mathrm{ac}})/(3\rho\omega a_i^2 R_{fl}^2)]\cos\omega t$ (in which $\boldsymbol{a}_i$ is the main axis of the ellipsoidal trajectory,[14] where $i=1, 2, 3$), and $\mathbf{v}$, the velocity of the bubble, is obtained from Eq. (7).

(4) $\mathbf{F}_I=4/3\pi R(t)^3(d\mathbf{U}(t)/dt)$, inertial force which comes from the fact that Magnaudet and Legendre[15] derived the expression for the drag in the non-inertial frame translating with the bubble.

(5) $\mathbf{F}_{\mathrm{drag}}=-4\pi\rho\nu R(t)\mathbf{U}(t)$ (for the condition when both the translational and the radial Reynolds numbers are small) and $\mathbf{F}_{\mathrm{drag}}=-12\pi\rho\nu R(t)\mathbf{U}(t)$ (for the condition when at least one of the Reynolds numbers is large) are the viscous drag forces on the moving SL bubble.

(6) $\mathbf{F}_H=8\pi\rho\nu\int_0^t\exp[9\nu\int_\tau^t R(t')^{-2}dt']\mathrm{erfc}[\sqrt{9\nu\int_\tau^t R(t')^{-2}dt'}]\times(d[R(\tau)\mathbf{U}(\tau)]/d\tau)d\tau$ is the history force which is produced due to the wake behind the bubble. In order to reduce the kernel to only one variable, we perform the same manipulations done by Toegel *et al.*[14] Considering the effect of the history force on the bubble translational-radial dynamics, the exponential term is replaced with a suitable decay constant $\exp(H)\mathrm{erfc}(\sqrt{H})\approx\exp(-\alpha H)$. With this substitution, the approximate kernel can be written as a product of factors which depend on one variable only, i.e., either on t or τ. The differential translational-radial dynamics of the bubble [Eq. (7)] is obtained by using dimensionless time: $H(t) := 9\nu\int_{-\infty}^{\tau}R(t')^{-2}dt'$.

III. NUMERICAL RESULTS AND DISCUSSION

Figures 1–5 show the calculated results for a moving SL bubble in *N*-methylformamide. To solve the coupled translational-radial equation of the moving SL bubble [Eq. (7)], we used the fourth-order Runge–Kutta algorithm. Based on the reported experimental work,[7] the calculations are carried to the liquid temperature of 18 °C and driving frequency of 30 kHz. So we selected the parameters as $P_o = 101\,325$ Pa, $\sigma = 0.038$ N m^{-1}, $\nu = 1.65$ m^2 s^{-1}, $\rho = 1000$ kg/m^3, and $c = 1660$ m/s. We examined the bubble movement by considering the diffusive stability requirement[31] for the amplitude of the driving pressure in the domain of Pa = 1.4–1.65 atm. At pressures smaller than Pa = 1.36 atm and higher than Pa = 1.67 atm, the calculations did not show stable trajectory. This is in good agreement with the experimental report in which the Pa = 1.6 atm and Pa = 1.3 atm are introduced as up and down scale amplitudes of the driving pressures, respectively. The m-SBSL has also been observed in *N*-methylformamide.[7] The ambient radius of the bubble is selected in the range of $R_o = 7.0$–9.5 μm according to the diffusive equilibrium stability[31] for a SL bubble.

Figure 1 shows the bubble trajectory at various driving pressures and equilibrium radii. In the numerical calculation, we have found that in order to obtain a stable trajectory for the moving bubble, the bubbles should be allowed to start their translational motion near the origin of the resonator, namely, sub-millimeter distance for *X*- and *Y*-directions and millimeter distance for the *Z*-direction. We chose the point of $X_o = 0.075$ mm, $Y_o = -0.17$ mm, and $Z_o = 1.24$ mm in this work. For the small deviations of the starting point from the origin in the above mentioned range, we have some oscillations in the first several hundred cycles. Finally, we obtained

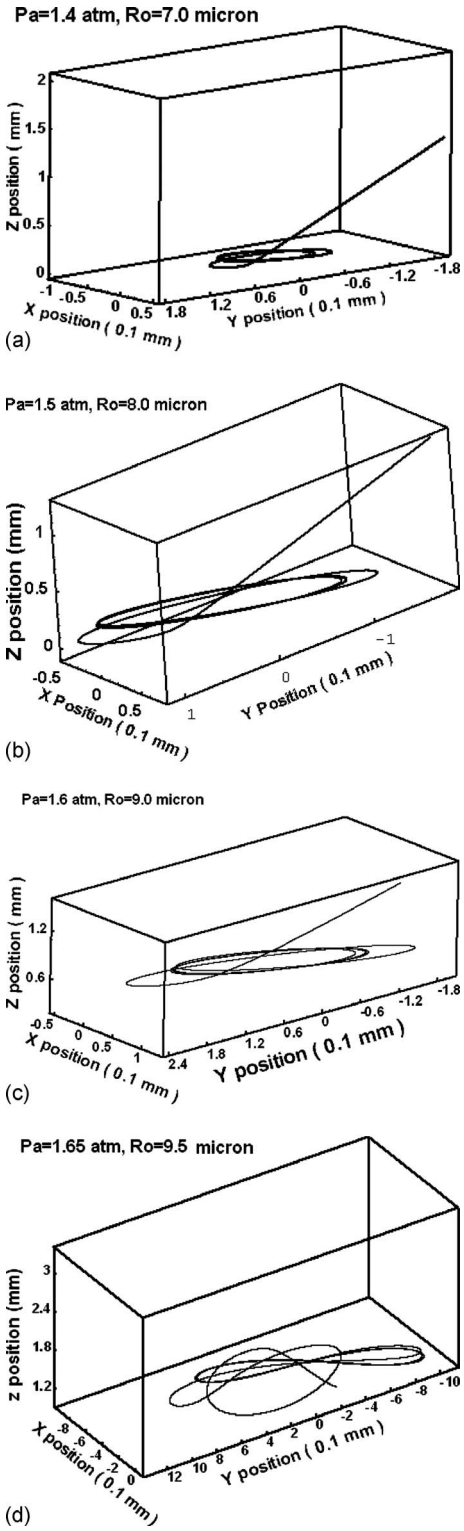


FIG. 1. Numerically calculated trajectories of a SL bubble in *N*-methylformamide. Calculations have been done for the amplitude of driving pressures in the range Pa=1.4–1.65 atm in which the experimental observation is reported (Ref. 7). Frequency of the acoustic field is set equal to $f=30$ kHz. Bubble ambient radii increase with the amplitude of the driving pressure to satisfy the diffusive equilibrium stability conditions (Ref. 31): (a) Pa=1.4 atm and $R_o=7.0$ μm, (b) Pa=1.5 atm and $R_o=8.0$ μm, (c) Pa=1.6 atm and $R_o=8.0$ μm, and (d) Pa=1.65 atm and $R_o=9.5$ μm. For higher driving pressures, bigger trajectories have been calculated. At the *Z*-direction due to the buoyancy force, the bubbles oscillate around a point above the central antinodes.

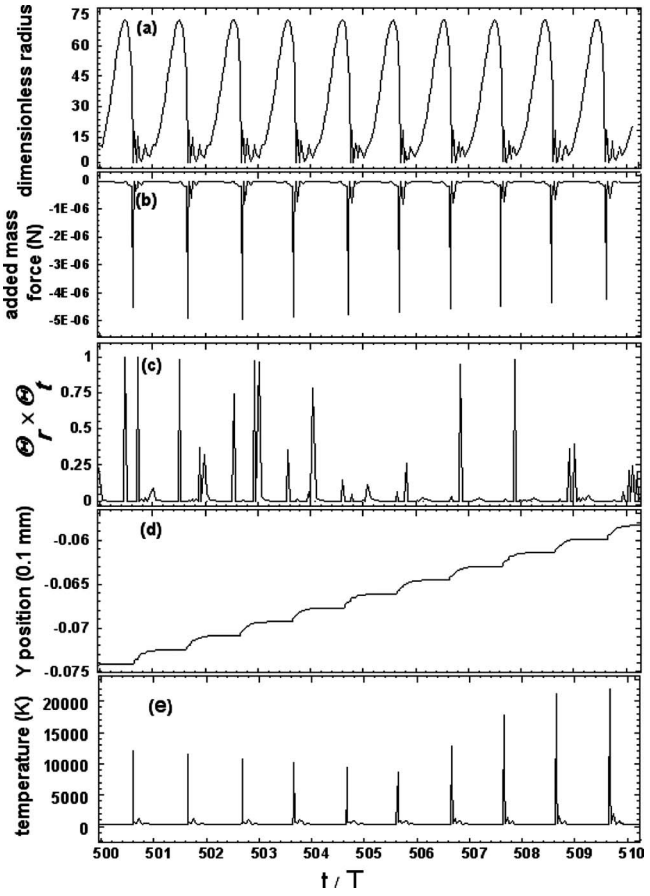


FIG. 2. Calculated properties of the typical moving bubble with Pa =1.5 atm and $R_o=8.0$ μm for ten acoustic cycles. (a) The repetitious bubble radial oscillations, (b) time dependent variations of the added mass force on the bubble, (c) the region of appearance of history force, (d) the temporal displacement of bubble in *Y*-direction, and (e) the bubble peak temperature.

the stable bubble trajectory motion. Therefore, the selected starting point does not affect the obtained results. As shown in Fig. 1, in the *Z*-direction, the effect of the buoyancy force causes the bubble to move around a point which lies above the central antinodes of the acoustic field. It should be clarified that in SBSL condition, the balance between $\mathbf{F}_{\mathrm{Bj}}$ and $\mathbf{F}_{\mathrm{buoy}}$ identifies the bubble levitation position near the central antinodes of the standing sound field in the liquid.[24,32] By increasing the driving pressure, the bubble moves in the trajectory with larger domain. An explanation of the dependency of the bubble trajectory on the driving pressure requires the consideration of the instantaneous translational-radial dynamics of the bubble during the acoustic cycles.

Figure 2 shows the temporal evolution of the coupled translational-radial dynamics and the effects of both the bubble translational and radial motions on the bubble peak temperature during ten acoustic cycles, typically for the same bubble in Fig. 1(b). Figure 2(a) shows the regular oscillations of an acoustically driven bubble. The dependence of the components of hydrodynamic force on the variations of the bubble radius and the flow of the fluid around the bubble, induce the required momentum of the bubble's translational motion in each acoustic cycle. In Fig. 2(b), the added mass force on the moving SL bubble has been shown. At the

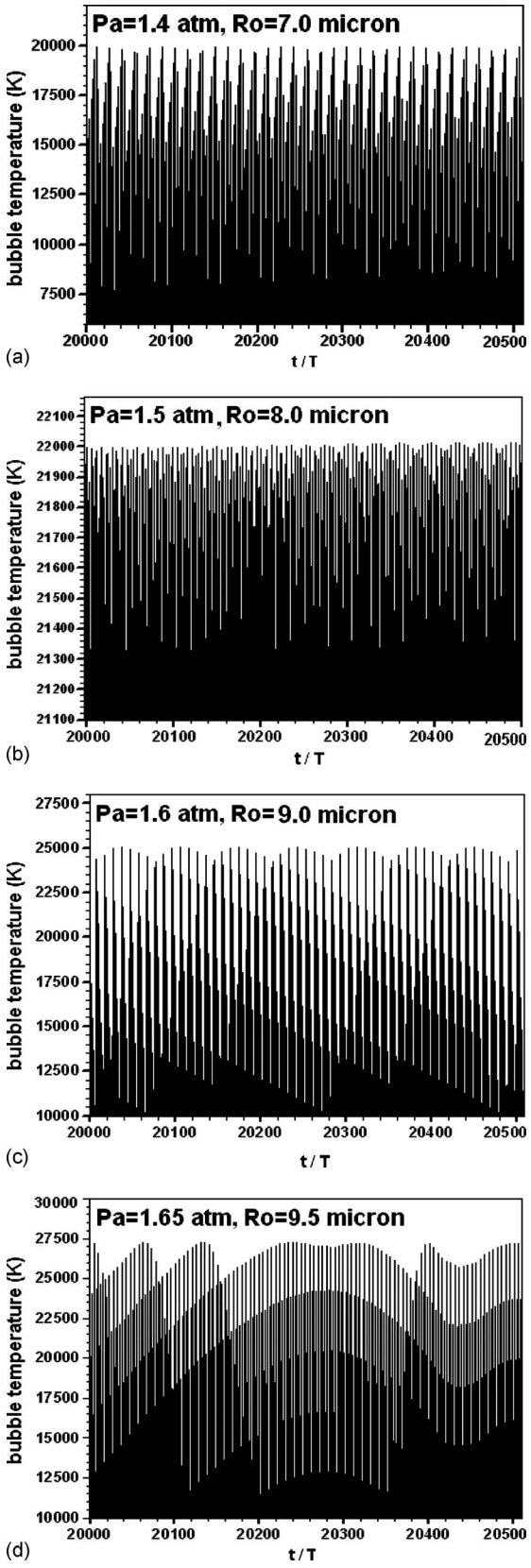


FIG. 3. Peak temperatures of the moving SL bubbles of Fig. 1, in 500 (20 000–20 500) acoustic cycles. Coupled radial-translational dynamics of bubble effects on the bubble size and its wall velocity and consequently on the strength of the bubble collapse. Due to the different compression ratios of the moving bubble, different peak temperatures have been calculated. Similar to the SBSL condition, the bubble peak temperature increases with the increase in the amplitude pressure.

bubble collapse, the rapid increase in the velocity of the displaced fluid leads to the increase in the added mass force. In Fig. 2(c), the instant of the appearance of the history force on the bubble has been shown by the product of $\Theta_r \times \Theta_t$. The effect of the history force on the dynamics of moving SL bubble vanishes when the products of $\Theta_r = 1/(1+(\mathrm{Re}_r(t)/\mathrm{Re}_{r,\mathrm{crit}}(t))^4$ and $\Theta_t = 1/(1+(\mathrm{Re}_t(t)/\mathrm{Re}_{t,\mathrm{crit}}(t))^4$ equal to zero. Due to the dependency of the radial and translational Reynolds numbers on the radial and the translational velocities, we could expect a zero history force at the instant of the bubble collapse. Figure 2(c) also shows a nonregular effect of the history force on the bubble. Because of the coupled radial-translational dynamics of the bubble, neither the radial Reynolds number $\mathrm{Re}_r = R|\dot{R}|/\nu$ nor the translational Reynolds number $\mathrm{Re}_t = R|\mathbf{u}-\mathbf{v}|/\nu$ has normal variations. This results in having different products of Θ_t and Θ_r during the different acoustic cycles. Our calculations show, during most of the acoustic cycles, that the bubble experiences no history force at the collapse. The magnitudes of the $\mathbf{F}_{\mathrm{Bj}}$ and $\mathbf{F}_{\mathrm{buoy}}$ can also be neglected because of the bubble's very small size. The effect of the quasi-steady viscous drag force $[\mathbf{F}_{\mathrm{drag}} = -4\pi\rho\nu R(t)\mathbf{U}(t)]$ can be neglected compared to the added mass force at the instant of the bubble collapse. Although the history force is the origin of generating the bubble trajectory,[14] based on the above discussion, the added mass force is the origin of the shift in the bubble translational motion at the instant of its collapse [Fig. 2(d)]. The effect of the coupled translational-radial dynamics of the bubble on its peak temperature is shown in Fig. 2(e). The size of the moving bubble varies because of its varying height position (z-axis) in the trajectory. Due to the different bubble sizes at its expansion phase, different strengths of its collapse are produced. Different strengths of the compression heating lead to different bubble peak temperatures.

In Fig. 3, we show the variation in the bubble peak temperatures in the moving sonoluminescence condition for the same bubbles in Fig. 1. This shows the superiority of the quasi-adiabatic model used here for the temperature prediction in comparison to the earlier isothermal model.[14] In contradiction to the constant bubble peak temperature in SBSL condition, the peak temperature of a moving SL bubble changes because of the coupled radial-translational dynamics. Translational motion of the bubble causes the bubble to get different equilibrium radii in different positions in liquid. Therefore, due to the different compression ratios, different peak temperatures appear when the bubble collapse takes place. Here, similar to the SBSL condition, we see the increase in the bubble peak temperature with the increase in the driving pressure.

Figure 4 shows the components of the bubble trajectories of bubbles in Fig. 1. In the figure, we see the effect of the increasing amplitude of the driving pressure on the shape of the bubble trajectory. The increase in the amplitude of the driving pressure increases the domain of the trajectory, which results in decreasing the number of trajectory circulations around the point close to the central antinodes. To show the effect of the dominant force on the bubble collapse which

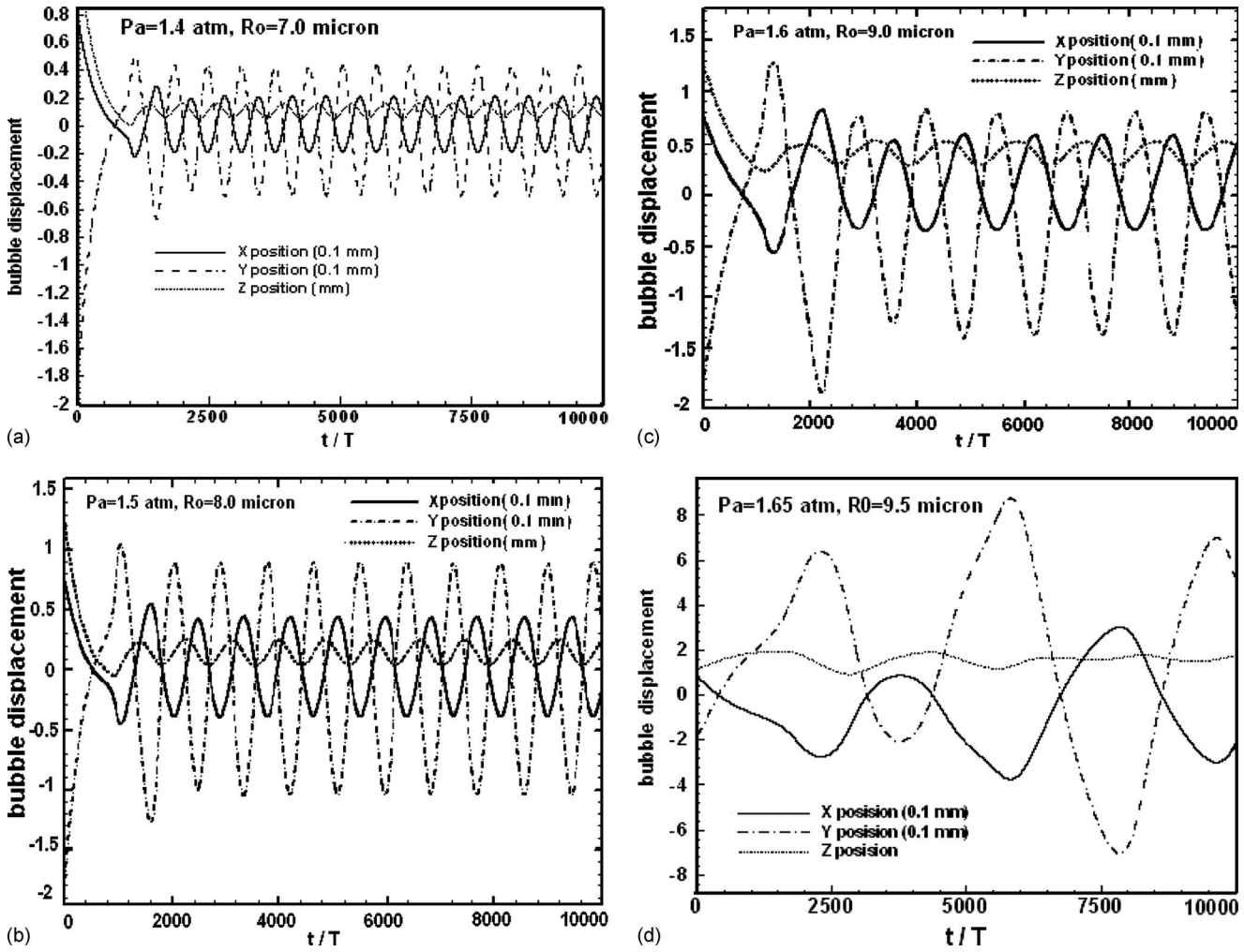


FIG. 4. Components of the trajectory of the same bubbles in Fig. 1. With increasing amplitude of the driving pressure, the domain of the bubble trajectory increases and the number of oscillations around the central antinodes of the acoustic standing field in *N*-methylformamide decreases in the same time duration.

causes acceleration on the bubble displacement, we calculated the added mass force maxima on the bubble for various driving pressures.

In Fig. 5, the added mass force maxima for the driven bubbles in Fig. 1 have been shown. It is apparent that the increase in the average added mass force maxima causes the longer trajectory path and the duration for the moving bubble around the central antinodes of the sound field. From Fig. 5, we find that the driven bubble in lower pressure experiences the smaller acceleration and smaller added momentum at the instant of its collapse. This leads to the smaller domain of the bubble trajectory compared to that in the higher driving pressures.

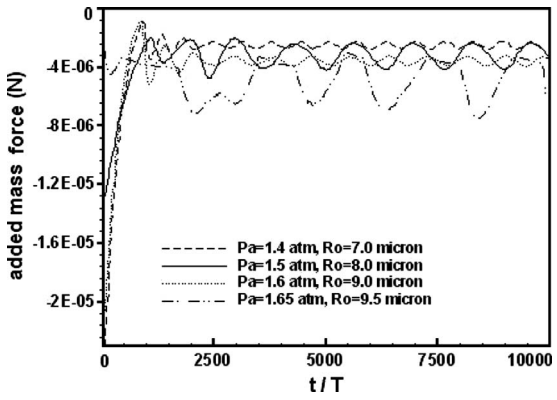


FIG. 5. Comparison of the added mass force maxima for the same bubbles in Fig. 1. The magnitude of added mass force on the bubble is proportional to the bubble size and its wall velocity. At higher driving pressures, the more intense bubble collapse leads to larger amounts of added mass force and consequently higher acceleration on bubble translational. The result is the larger amplitude of the bubble trajectory at the higher driving pressures.

IV. CONCLUSIONS

In this paper, we investigated the dynamics of a moving SL bubble from a different point of view. In our model, we used an adiabatic system, instead of the isothermal model, which is not suitable for sonoluminescence condition, especially at high driving pressures. The peak temperature of the moving bubble is calculated and its dependency on the bubble translational motion is shown. Translational motion of the bubble affects the size of the bubble because of bubble movement in different heights. In our calculation, different peak temperatures are observed due to different strengths of the collapse. We showed that the influence of the buoyancy force causes the translational movement of the SL bubble around a point above the central antinode of the sound field.

At the bubble collapse, added mass force dominates the other components of the hydrodynamic force on the bubble. Because of the increase in the added mass force, a shift in the bubble translational movement takes place. The size of the bubble trajectory depends on the amplitude of the driving pressure. Higher driving pressure leads to trajectory with bigger domain and longer period.

ACKNOWLEDGMENTS

We gratefully acknowledge Professor K. Suslick for his constructive comments and approving of Fig. 4 based on his valuable experimental observations, and Dr. E. Lotfi. We also acknowledge the Sharif University of Technology for financial support of the project.

[1]D. F. Gaitan, L. A. Crum, C. C. Church, and R. A. Roy, "Sonoluminescence and bubble dynamics for a single, stable, cavitation bubble," J. Acoust. Soc. Am. **91**, 3166–3188 (1992).
[2]I. Akhatov, R. Mettin, C. D. Ohl, U. Arlitz, and W. Lauterborn, "Bjerknes force threshold for stable single bubble sonoluminescence," Phys. Rev. E **55**, 3747–3750 (1997).
[3]A. I. Eller, "Force on a bubble in a standing acoustic wave," J. Acoust. Soc. Am. **43**, 170–171 (1968).
[4]L. A. Crum, "Bjerknes forces on bubbles in stationary sound field," J. Acoust. Soc. Am. **57**, 1363–1370 (1975).
[5]A. Moshaii and R. Sadighi-Bonabi, "Role of liquid compressional viscosity in the dynamics of a sonoluminescing bubble," Phys. Rev. E **70**, 016304 (2004).
[6]A. Moshaii, R. Sadighi-Bonabi, and M. T. Rahini, "Effects of bulk viscosity in non-linear bubble dynamics," J. Phys.: Condens. Matter **16**, 1687–1694 (2004).
[7]Y. T. Didenko, W. B. McNamara, and K. S. Suslick, "Molecular emission from singlebubble sonoluminescence," Nature (London) **407**, 877–897 (2000).
[8]S. D. Hopkins, S. J. Putterman, B. A. Kappus, K. S. Suslick, and C. G. Camara, "Dynamics of a sonoluminescing bubble in sulfuric acid," Phys. Rev. Lett. **95**, 254301 (2005).
[9]F. B. Seeley, D. A. Gregory, S. Thampson, and J. D. Brown, "Temporally stabilized sonoluminescence in ethylene glycol," ARLO **6**, 48–52 (2005).
[10]D. J. Flannigan and K. S. Suslick, "Formation and temperature measurement during single-bubble cavitation," Nature (London) **434**, 52–55 (2005).
[11]D. J. Flannigan and K. S. Suslick, "Plasma line emission during single-bubble cavitation," Phys. Rev. Lett. **95**, 044301 (2005).
[12]D. J. Flannigan and K. S. Suslick, "Molecular and atomic emission during single-bubble cavitation in concentrated sulfuric acid," ARLO **6**, 157–161 (2005).
[13]A. Troia, D. M. Ripa, and R. Spagnolo, "Moving single bubble sonoluminescence in phosphoric acid and sulphuric acid solution," Ultrason. Sonochem. **13**, 278–282 (2006).
[14]R. Toegel, S. Lutter, and D. Lohse, "Viscosity destabilizes sonoluminescing bubbles," Phys. Rev. Lett. **96**, 114301 (2006).
[15]J. Magnaudet and D. Legendre, "The viscous drag force on a spherical bubble with a time-dependent radius," Phys. Fluids **10**, 550–554 (1998).
[16]A. J. Reddy and A. J. Szeri, "Coupled dynamics of translation and collapse of acoustically driven microbubbles," J. Acoust. Soc. Am. **112**, 1346–1352 (2002).
[17]S. Hilgenfeldt, S. Grossman, and D. Lohse, "Sonoluminescence light emission," Phys. Fluids **11**, 1318–1330 (1999).
[18]K. Yasui, "Alternative model of single bubble sonoluminescence," Phys. Rev. E **56**, 6750–6760 (1997).
[19]B. D. Storey and A. Szeri, "Water vapor, sonoluminescence and sonochemistry," Proc. R. Soc. London, Ser. A **457**, 1675–1700 (2000).
[20]R. Toegle and D. Lohse, "Phase diagrams for sonoluminescing bubbles: A comparison between experiment and theory," J. Chem. Phys. **118**, 1863–1875 (2003).
[21]A. Moshaii, R. Rezaei-Nasirabad, Kh. Imani, M. Silatani, and R. Sadighi-Bonabi, "Role of thermal conduction in single bubble cavitation," Phys. Lett. A **372**, 1283–1287 (2008).
[22]D. Legendre, J. Boree, and J. Magnaudet, "Thermal and dynamic evolution of a spherical bubble moving steadily in a superheated or subcooled liquid," Phys. Fluids **10**, 1256–1272 (1998).
[23]U. Parlitz, R. Mettin, S. Luther, I. Akhatov, M. Voss, and W. Lauterborn, "Spatio-temporal dynamics of acoustic cavitation bubble clouds," Philos. Trans. R. Soc. London, Ser. A **357**, 313–334 (1999).
[24]T. J. Matula, "Bubble levitation and translation under single-bubble sonoluminescence conditions," J. Acoust. Soc. Am. **114**, 775–781 (2003).
[25]Z. Galavani, R. Rezaei-Nasirabad, and R. Sadghi-Bonabi, "Hydrodynamic force on acoustically driven bubble in sulfuric acid," July 2008, *Proceedings of World Academy of Science, Engineering and Technology* **43**, 560–564, Heidelberg, Germany.
[26]J. B. Keller and M. J. Miksis, "Bubble oscillations of large amplitude," J. Acoust. Soc. Am. **68**, 628 (1980).
[27]R. Lofstedt, B. P. Barber, and S. J. Putterman, "Toward a hydrodynamic theory of sonoluminescence," Phys. Fluids A **5**, 2911–2918 (1993).
[28]A. Prosperetti, "Thermal effects and damping mechanisms in the forced radial oscillations of gas bubbles in liquid," J. Acoust. Soc. Am. **61**, 17–27 (1977).
[29]J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (Wiley, New York, 1954).
[30]G. Stokes, "On the effect of internal friction of fluids on the motion of pendulum," Trans. Cambridge Philos. Soc. **9**, 8–27 (1851).
[31]M. M. Fyrillas and A. J. Sezri, "Dissolusion or growth of soluble spherical bubbles," J. Fluid Mech. **277**, 381–407 (1994).
[32]T. J. Matula, V. J. Bezzerides, P. R. Hilmo, L. N. Couret, T. W. Olson, L. A. Crum, J. E. Swalwell, D. W. Kuhns, and R. A. Roy, "The effect of buoyancy on sonoluminescing bubbles," ARLO **1**, 13–18 (2000).

The pulse tube and the pendulum

G. W. Swift and S. Backhaus
Condensed Matter and Thermal Physics Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 2 June 2009; revised 17 August 2009; accepted 20 August 2009)

An inverted pulse tube in which gravity-driven convection is suppressed by acoustic oscillations is analogous to an inverted pendulum that is stabilized by high-frequency vibration of its pivot point. Gravity acts on the gas density gradient arising from the end-to-end temperature gradient in the pulse tube, exerting a force proportional to that density gradient, tending to cause convection when the pulse tube is inverted. Meanwhile, a nonlinear effect exerts an opposing force proportional to the square of any part of the density gradient that is not parallel to the oscillation direction. Experiments show that convection is suppressed when the pulse-tube convection number $N_{\mathrm{ptc}} = \omega^2 a^2 \sqrt{\Delta T / T_{\mathrm{avg}}} / [g(\alpha D \sin\theta - L\cos\theta)]$ is greater than 1 in slender tubes, where ω is the radian frequency of the oscillations, a is their amplitude, ΔT is the end-to-end temperature difference, T_{avg} is the average absolute temperature, g is the acceleration of gravity, L is the length of the pulse tube and D is its diameter, α is about 1.5, and the tip angle θ ranges from 90° for a horizontal tube to 180° for an inverted tube. Theory suggests that the temperature dependence should be $\Delta T / T_{\mathrm{avg}}$ instead of $\sqrt{\Delta T / T_{\mathrm{avg}}}$. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3238156]

PACS number(s): 43.35.Ud, 43.25.Ts, 43.28.Py [RR] Pages: 2273–2284

I. INTRODUCTION

Rigid pendula exhibit many interesting phenomena,[1] including dynamic stabilization: If the pivot point of a rigid pendulum is vibrated at high enough frequency and high enough amplitude, the pendulum tends to align with the vibration axis, and the pendulum can stand inverted, seeming to defy gravity. The equation of motion can be derived with the simple approach of Blitzer.[2] Let the x and y directions, the angle θ of the pivot-point vibration relative to gravity g, and the angle $\phi(t)$ of the pendulum relative to gravity be defined as shown in Fig. 1(a), and let m be the mass of the pendulum's bob and L be the length of its rod. The positions $x(t)$ and $y(t)$ of the pendulum's bob are given by

$$x = a\sin\theta\cos\omega t + L\sin\phi, \tag{1}$$

$$y = -a\cos\theta\cos\omega t - L\cos\phi, \tag{2}$$

when the pivot point is forced to vibrate sinusoidally with amplitude a and radian frequency ω. The equation of motion for the bob can be written as

$$F_x = m\ddot{x}, \quad F_y - mg = m\ddot{y}, \tag{3}$$

where the F's are the two components of force exerted by the rod on the bob and the overdots represent time derivatives. By Newton's third law, the bob exerts force $-\vec{F}$ on the rod, applied at the end of the rod attached to the bob. The torque on the rod about the pivot point due to $-\vec{F}$ must be zero, because this massless rod has zero moment of inertia, I_{rod}. Thus

$$-LF_x\cos\phi - LF_y\sin\phi = I_{\mathrm{rod}}\ddot{\phi} = 0. \tag{4}$$

Combining these four equations by eliminating x, y, F_x, and F_y yields the equation of motion for the pendulum angle ϕ as follows:

$$\ddot{\phi} = -(g/L)\sin\phi + \omega^2(a/L)\sin(\theta - \phi)\cos\omega t. \tag{5}$$

For $0 < \phi < 180°$, the torque applied by gravity that tends to decrease ϕ is apparent in Fig. 1(a) and Eq. (5). The time-averaged torque caused by the pivot-point vibration, tending to align ϕ with the vibration, is not so apparent in Fig. 1(a) or Eq. (5), but Fig. 1(b) helps explain the mechanism, if the vibration is exaggerated and gravity is neglected.[1] The figure shows the pendulum at two extremes j and k of its motion under this exaggerated circumstance. At j, the acceleration of the pivot point causes a large, positive $\dot{\phi}$, without much acceleration of the bob. At k, the acceleration of the pivot point causes the bob to accelerate parallel to the vibration direction, with only a small, negative $\dot{\phi}$. The net effect on ϕ is positive, causing the pendulum to tend to align with the vibration. In other words, the time-averaged torque tending to align ϕ with the vibration is proportional to the product of the amplitude of the angular vibration, $\phi_k - \phi_j$, and how strongly the torque caused by the vibration varies with ϕ.

For decades, quantitative analysis of Eq. (5) when $\omega^2 \gg g/L$ has appeared as an exercise in textbooks on classical mechanics, such as Ref. 3 for $\theta = 180°$ and $\theta = 90°$ and Ref. 4 for $\theta = 180°$. Extended to arbitrary θ, the analysis shows that the dimensionless number

$$N_{\mathrm{pendulum}} = \frac{\omega^2 a^2}{gL} \tag{6}$$

determines the simple pendulum's behavior for slow time scales, i.e., time scales $\gg 1/\omega$. For $N_{\mathrm{pendulum}} > 4$, the pendulum can be stably held up for any θ. For $1 < N_{\mathrm{pendulum}} < 4$, it can be stably held up only if the pivot-point's vibration is close enough to vertical (i.e., close enough to either $\theta = 0°$ or $\theta = 180°$, which are equivalent). For $N_{\mathrm{pendulum}} < 1$, the pendulum cannot be stably inverted for any θ. (The details of this

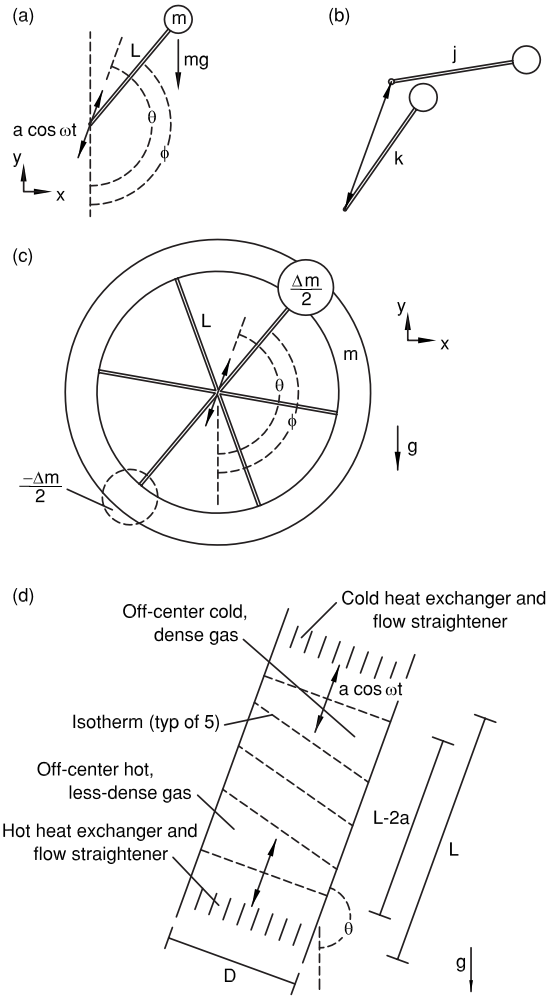


FIG. 1. (a) An inverted, rigid pendulum whose support is vibrated at a high enough frequency and amplitude can be stabilized against falling over. Relative to gravity, which points in the $-y$ direction, θ gives the angle along which the vibrations occur, and $\phi(t)$ gives the angle of the pendulum. (b) Consideration of an exaggerated situation, in which the amplitude of the pivot-point oscillation is not small compared with the length of the pendulum's rod, illustrates the stabilization mechanism. (c) An out-of-balance ring that is vibrated at its central pivot point behaves similarly, due to the same mechanism, but described by slightly more complicated mathematics. (d) An inverted pulse tube whose gas is oscillated at a high enough frequency and amplitude can be stabilized against natural convection, because of a similar mechanism. The double-headed arrows show the peak-to-peak amplitude $2a$ of the acoustic oscillation, assumed to carry all isotherms. Ideally, the central slug of gas, whose length is $L-2a$, would experience no gravity-driven convection.

analysis are omitted here, because similar details are presented in the next paragraph and in Sec. II.)

A more complicated pendulum sets the stage well for analysis of the pulse tube in Sec. II. Figure 1(c) shows a ring of radius L and mass m, supported from its central pivot point by massless spokes. The mass m is uniformly distributed around the ring, except for an out-of-balance part $\Delta m/2$, which has been removed from location $-\phi$ and added to location $+\phi$. The pivot point is forced to vibrate sinusoidally along a line at angle θ from the vertical, with amplitude a and angular frequency ω, as for the simple pendulum considered above, and the analysis again begins by following Blitzer's approach.[2] The coordinates of the "positive" mass at $+\phi$ and the "negative" mass at $-\phi$ are written down, as are equations of motion for these two masses. The torque exerted on the ring by its contact with the two out-of-balance masses is set equal to $I_{\text{ring}}\ddot{\phi}$, where $I_{\text{ring}}=mL^2$ is the ring's moment of inertia. Eliminating the forces and coordinates yields an equation of motion for the angle $\phi(t)$,

$$mL^2\ddot{\phi} = -\Delta mgL\sin\phi + \Delta m\omega^2 aL\sin(\theta-\phi)\cos\omega t \quad (7)$$

$$\equiv \tau_{\text{grav}} + \tau_{\text{vib}}, \quad (8)$$

where the right-hand side can be interpreted as the sum of a torque due to gravity and a torque due to the vibration of the pivot. References 3 and 4 show that the pendulum angle ϕ should be written as the sum of slow and fast parts

$$\phi = \phi_{\text{slow}} + \phi_{\text{fast}}\cos\omega t, \quad (9)$$

when $\omega^2 \gg \Delta mg/mL$. Using Taylor-series expansions

$$\tau = \tau|_{\phi_{\text{slow}}} + \left.\frac{\partial\tau}{\partial\phi}\right|_{\phi_{\text{slow}}}\phi_{\text{fast}}\cos\omega t \quad (10)$$

for both terms on the right-hand side of Eq. (7) and time averaging confirms that there was no need for a sin ωt term in Eq. (9) and it generates two equations: a "fast" equation from the cos ωt terms and a "slow" equation from terms with nonzero time-averages,

$$\phi_{\text{fast}} = -\frac{\Delta m}{m}\frac{a}{L}\sin(\theta-\phi_{\text{slow}}), \quad (11)$$

$$mL^2\ddot{\phi}_{\text{slow}} = -\Delta mgL\sin\phi_{\text{slow}} + \frac{(\omega a\Delta m)^2}{4m}\sin(2\theta-2\phi_{\text{slow}}). \quad (12)$$

If this pendulum can be stably inverted, then $\ddot{\phi}_{\text{slow}}=0$ and setting the right-hand side of Eq. (12) equal to zero must yield a solution for ϕ_{slow}. Whether such a solution exists, and its value, depends on the ratio of the coefficients of the two terms on the right (dropping the factor of 4 for simplicity),

$$N_{\text{ring}} = \frac{\omega^2 a^2}{gL}\frac{\Delta m}{m}. \quad (13)$$

It is interesting how many effects combine to put Δm and $(\Delta m)^2/m$ in the two terms on the right-hand side of Eq. (12), and hence to put $\Delta m/m$ in Eq. (13). The left-hand side of Eq. (12) is the moment of inertia times the slow angular acceleration, so the right-hand side must be the slow torque. The first term on the right-hand side shows gravity applying torque proportional to the unbalanced mass Δm, with the mass $\Delta m/2$ trying to drop down on the right and the negative mass $-\Delta m/2$ trying to float up on the left, both applying clockwise torques in Fig. 1(c). The second term on the right-hand side, which represents the time-averaged torque tending to align the unbalanced axis of the ring with θ, arises from the mechanism illustrated in Fig. 1(b), i.e., the time-averaged product of $\partial\tau_{\text{vib}}/\partial\phi$ and ϕ_{fast} in Eq. (10), where τ_{vib} is the vibration torque given by the second term in Eq. (7). The unbalanced mass Δm appears in the numerator of ϕ_{fast} in Eq. (11) because it is responsible for τ_{vib} through the mechanism shown in Fig. 1(b). The ring mass m appears in the denomi-

nator of ϕ_{fast} in Eq. (11) because the ring's moment of inertia resists the fast torque. Finally, $\partial\tau_{\text{vib}}/\partial\phi$ retains the mass dependence of τ_{vib} itself, namely, Δm. Overall, N_{ring} represents the angle-independent part of the ratio of the time-averaged alignment torque [proportional to $(\Delta m)^2/m$] to the gravitational torque (proportional to Δm).

This paper explores the hypothesis that a similar mechanism is responsible for suppressing natural convection in the gas in a tube with an axial temperature gradient, when the gas oscillates axially at high enough frequency and amplitude. As shown in Fig. 1(d) for such a tube with its cold end higher than its hot end, gravity tends to pull the dense gas near the cold end to one side and down, pushing the less-dense gas near the hot end to the other side and up. This puts the center of mass of the gas below the tube's centerline. The vibration can then exert a time-averaged torque on the entire body of gas, via time-averaged oscillating forces on this off-center center of mass, in a process analogous to that shown in Fig. 1(b). This torque opposes that of gravity and can balance it, preventing convection.

The gas in a pulse tube experiences such axial oscillations and supports such an axial temperature gradient. The pulse tube, a vital component of cryogenic pulse-tube refrigerators,[5] is a smooth-walled tube without internal structures, bounded on both ends by flow straighteners and heat exchangers through which gas flows easily. Its purpose is to transmit acoustic power through the gas from the cold end to the hot end with minimal heat leak from the hot end to the cold end. The lack of internal structure generally makes low-loss transmission of acoustic power easy, but makes heat-leak minimization challenging. The peak-to-peak volumetric stroke of the moving gas is always less than the volume of the gas in a pulse tube. Ideally, one imagines a perfectly thermally stratified slug of gas, whose volume is the difference between the tube volume and the volumetric stroke, oscillating axially in the tube and remaining entirely inside the tube at all times, conducting only a little heat from hot to cold, without any accompanying convection. However, several heat-transfer mechanisms can disturb this ideal picture, carrying much more heat than would be carried by conduction alone, unless attention is paid to minimizing each one of them. One such heat-transfer mechanism—natural convection due to gravity acting on density gradients in the gas, the subject of this paper—is known to occur commonly in low-frequency pulse-tube refrigerators, but it is also known that such convection is often reduced or absent in high-frequency pulse-tube refrigerators.[6–8]

Our motivation for this work arose in the context of pulse-tube refrigerators and thermoacoustic engines, sometimes coupled, in which convectively stable orientation of the tubes relative to gravity was inconvenient and an accurate understanding of the suppression of convection by high-frequency oscillations was desired. In thermoacoustic-Stirling hybrid engines[9] and cascade thermoacoustic engines,[10] the tubes that transmit acoustic power across a temperature difference while minimizing heat leak are called thermal-buffer tubes. They generally carry acoustic power from a hot temperature to ambient temperature, while pulse tubes carry acoustic power from a cold temperature to ambient temperature. But even in pulse-tube refrigerators, these tubes are sometimes called thermal-buffer tubes. For brevity in this paper, all such tubes are referred to as pulse tubes, and their end temperatures are labeled as hot and cold.

Below, theoretical arguments (Sec. II) and experimental evidence (Sec. III) are presented to show that

$$N_{\text{ptc}} = \frac{\omega^2 a^2}{g(\alpha D \sin\theta - L\cos\theta)}\left(\frac{\Delta T}{T_{\text{avg}}}\right)^{\beta} \tag{14}$$

is a useful and plausible choice of dimensionless group for characterizing this phenomenon in pulse tubes of low-aspect ratio D/L. As above, ω is the radian frequency of the oscillation, a is its displacement amplitude, and g is the acceleration of gravity; D is the pulse-tube diameter and L is its length, ΔT is the end-to-end temperature difference, and T_{avg} is the average temperature. The tip angle θ is taken to be zero in the vertical, gravitationally stable orientation, and this equation is only valid for $90° \le \theta \le 180°$ (where $\cos\theta \le 0$, so both terms in the denominator are non-negative). The parameter α is a fitting parameter discussed below, experimentally found to be about 1.5, and experiment shows that β is close to 1/2 while theory suggests $\beta = 1$.

II. THEORY

An extensive literature describes the interaction between rapid vibration and steady convection in fluids, in the framework of the Boussinesq approximation, namely, that density variations due to temperature variations are small and density variations due to pressure variations are zero. This literature is reviewed and its foundations are succinctly summarized by Gershuni and Lyubimov.[11] After writing the hydrodynamic and thermal variables as the sum of fast variations at the vibration frequency and slow variations, they derive time-averaged equations of motion for the slow variables similar in spirit to Eq. (12) above, showing that fast vibrations effectively add a time-averaged body force to the fluid, whose magnitude and direction depend on the magnitude and direction of the vibration velocity and the fluid's temperature-gradient vector field. For steady state with negligible convection, their Eqs. (1.100) and (1.101) give the conditions for balance between gravity- and vibration-induced forces

$$\text{Ra}\vec{\nabla}T \times \hat{g} + \text{Ra}_{\text{vib}}\vec{\nabla}(\vec{w}\cdot\hat{n}) \times \vec{\nabla}T = 0, \tag{15}$$

$$\vec{\nabla}\cdot\vec{w} = 0, \tag{16}$$

$$\vec{\nabla}\times\vec{w} = \vec{\nabla}T \times \hat{n}, \tag{17}$$

where $\hat{g}$ and $\hat{n}$ are unit vectors in the directions of gravity and the vibration, respectively, T is the time-averaged temperature, $\vec{w}$ is the solenoidal part of $T\hat{n}$, and the ordinary Rayleigh number Ra and the vibrational Rayleigh number Ra_{vib} are given by

$$\text{Ra} = \frac{l^3 g \Delta T_{\text{char}}}{\nu\kappa T_{\text{char}}}, \tag{18}$$

$$\mathrm{Ra}_{\mathrm{vib}} = \frac{(\omega a l \Delta T_{\mathrm{char}})^2}{2\nu\kappa T_{\mathrm{char}}^2}, \tag{19}$$

with l a characteristic length of the boundary of the fluid, ΔT_{char} a characteristic temperature difference, T_{char} a characteristic temperature, ν a characteristic kinematic viscosity, and κ a characteristic thermal diffusivity. In Eqs. (18) and (19), we have set the thermal expansion coefficient of Ref. 11 equal to $1/T$, as appropriate for an ideal gas. Evident from Eq. (15), the existence of a steady state without convection depends on the ratio of $\mathrm{Ra}_{\mathrm{vib}}$ and Ra,

$$\frac{\omega^2 a^2}{g l}\frac{\Delta T_{\mathrm{char}}}{T_{\mathrm{char}}}. \tag{20}$$

Although derived in the context of the Boussinesq approximation, which is not really applicable to pulse tubes, this expression suggests most of the functional dependences that are displayed in Eq. (14), most of which are confirmed in the experiments described below. Presumably, numerical analysis based on Eqs. (15)–(17) could show whether the pulse-tube's length L, its diameter D, or some combination of those variables is best used for the characteristic length l, and could find the tip-angle dependence of vibrational suppression of convection in a pulse tube.

A high vibrational Rayleigh number tends to align density gradients along the direction of vibration, whether or not gravity is involved. Thus, we expect that this phenomenon also mitigates the effect of jet-driven streaming due to imperfect flow straightening and the effect of Rayleigh streaming, on Earth in zero gravity, because both of these streaming phenomena create non-axial density gradients in pulse tubes.[12] However, since streaming grows more intense as ωa rises, the mitigation cannot be as abrupt a function of ωa as it is for gravity-driven convection. Nevertheless, at a given ΔT, the effect of streaming might be reduced significantly.

The rest of this section presents a very simple attempt to anticipate the best choice for l in Eq. (20) when the pulse-tube's length L is significantly greater than its diameter D, which is a common situation in pulse-tube refrigerators. Although the approximations used here might seem crude, we hope that they can correctly capture the dominant functional dependences on D and L.

Three characteristic times are well separated. For a typical sinusoidally driven pulse-tube refrigerator, $1/\omega \sim 0.003$ s. This is significantly faster than the time required for an appreciable change in convective motion, estimated from the ring-pendulum analysis to be of the order of $\sqrt{l/g} \sim 0.1$ s. This, in turn, is significantly faster than the diffusive thermal-relaxation time $l^2/\kappa \sim 30$ s. Thus, for rough estimates, it is plausible to assume that temperatures are essentially carried with the moving gas on the time scales of the gas motion and that the dynamical behavior of the gravity-vibration interaction in the gas is qualitatively similar to that of the ring pendulum.

Furthermore, since $\nu \sim \kappa$ in gases, the viscous relaxation time for l-scale distance is also ~ 30 s, so the viscous penetration depth $\sqrt{2\nu/\omega}$ is typically much smaller than l. The velocity of the developing steady flow might be of the order of $l/\sqrt{l/g}$, so the steady-flow Reynolds number might initially be roughly $(l^2/\nu)/\sqrt{l/g} \sim 300$, a regime in which inertial effects are important and two- and three-dimensional flows are often time dependent. The typical Rayleigh number given in Eq. (18) can be estimated as $(l^2/\nu)(l^2/\kappa) \times (\Delta T_{\mathrm{char}}/T_{\mathrm{char}})/(l/g) \sim 10^6 \Delta T_{\mathrm{char}}/T_{\mathrm{char}}$, so modest $\Delta T_{\mathrm{char}}/T_{\mathrm{char}}$ can cause significant convection. Similarly, the typical vibrational Rayleigh number in Eq. (19) can be estimated as $10^8 (a/l)^2 (\Delta T_{\mathrm{char}}/T_{\mathrm{char}})^2$, so values of a/l that are common in pulse tubes can make $\mathrm{Ra}_{\mathrm{vib}} \sim \mathrm{Ra}$.

To keep the analysis of the problem simple, we retain the Boussinesq approximation, treating the gas in the pulse tube as incompressible. Thus, the double-headed arrows in Fig. 1(d), illustrating the peak-to-peak stroke of the gas, are taken to be the same length at the two ends of the pulse tube. The isotherms in Fig. 1(d) are shown at an instant of time when the motion of the gas is at mid-stroke, e.g., when $\omega t = -\pi/2$. A quarter cycle later, $\cos \omega t = 1$ and the uppermost isotherm would have just touched the cold heat exchanger; another half cycle later, when $\cos \omega t = -1$, the lowermost isotherm would just touch the hot heat exchanger. The slug of gas between these two isotherms, which always remains inside the pulse tube, is the object of interest. It has a length $L - 2a$, which we might take to be the effective length for this problem. However, our experiments cannot resolve the small difference between this length and L itself, so for simplicity we use L in the rest of this derivation.

The uppermost isotherm has temperature T_C when it is momentarily in contact with the cold heat exchanger at that temperature, but the pressure-induced adiabatic heating and cooling that the gas experiences causes its average temperature to be $T_{C,\mathrm{avg}} = T_C[1 + (\gamma - 1) p_a \sin \beta / \gamma p_m]$, where γ is the ratio of isobaric to isochoric specific heats, p_a is the pressure amplitude, p_m is the mean pressure, and $\beta = \pi/2$ is the phase by which oscillating pressure leads oscillating velocity (positive velocity going from hot to cold).[13] The hot isotherm's temperature $T_{H,\mathrm{avg}}$ obeys a similar expression. Our experiments cannot resolve the effects of these small p_a-dependent temperature differences, so for simplicity we describe the temperatures of the slug of gas with $\Delta T = T_H - T_C$ and $T_{\mathrm{avg}} = (T_H + T_C)/2$ in the rest of this derivation, instead of similar but more complicated expressions with $T_{H,\mathrm{avg}}$ and $T_{C,\mathrm{avg}}$.

As shown in Fig. 2(a), imagine that motion within this slug of gas in the pulse tube can be modeled as plug flow in a loop of piping that vibrates along the θ direction and whose cross-sectional area is half of the cross-sectional area A of the pulse tube itself, so gas rising on the left half of the pulse tube in Fig. 1(d) is modeled as rising plug flow in the left leg of Fig. 2(a), and similarly down on the right. In this model, the convective motion in the pulse tube is represented by a single degree of freedom, measured by a time-dependent displacement $\delta(t)$. This displacement and the superimposed vibration carry the isotherms, because the thermal-relaxation time is so much slower than the times for these motions, as estimated above. Then 2δ is the measure of how far the isotherms in the right half of the loop are misaligned from those in the left half at any instant of time, with the sign of δ as shown in the figure. Ignoring end effects for small δ, and assuming that end-to-end temperature differences are small enough that the density ρ can be assumed to be essentially

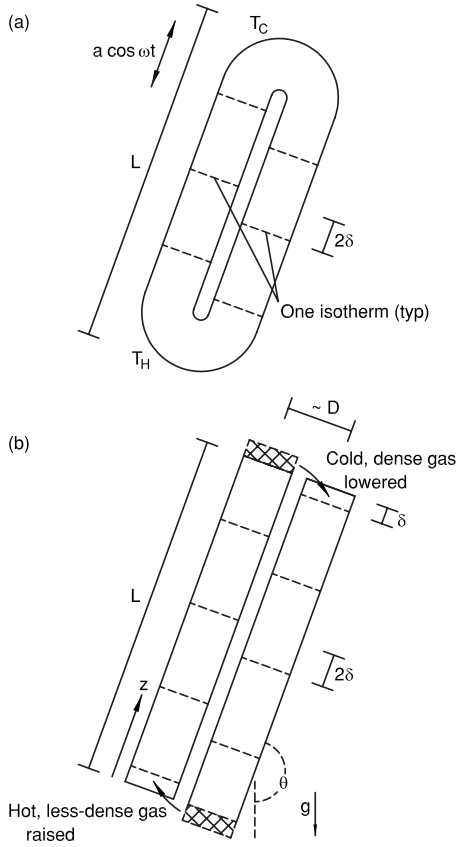


FIG. 2. (a) The convective flow in the pulse tube can be modeled crudely as plug flow in a loop of pipe, characterized by a single degree of freedom, measured by δ. As shown, the plug-flow displacement δ creates a misalignment of 2δ between isotherms on the left and right legs of the loop. A change in δ can be caused by gravity, or by the vibration acting on an off-center center of mass caused by nonzero δ itself. (b) The effect of a change in δ on the gravitational potential energy of the gas in such a loop of pipe can be estimated from an even more simplified model.

linear in position [not obviously a good assumption, but linear $T(z)$ is discussed below], the density in the two legs of the loop can be written as

$$\rho(z) = \rho_H + (z \pm \delta)\Delta\rho/L \tag{21}$$

except near the ends, where z is the distance from the hot end, $\Delta\rho = \rho_C - \rho_H$, the plus sign is chosen for the right leg and the minus sign for the left leg, and the subscripts on ρ correspond to those on T above. Thus, when $\delta = 0$, the density rises linearly from ρ_H at $z=0$ to ρ_C at $z=L$ in both legs, and nonzero δ shifts one of these density profiles up and the other one down.

A Lagrangian derivation of the equation of motion for $\delta(t)$ is well suited to keeping track of details here. The applied vibrational displacement $a\cos\omega t$ is superimposed on the plug-flow displacement δ, so the velocity of the gas is $-\omega a \sin\omega t + \dot{\delta}$ in the left leg of the loop and $-\omega a \sin\omega t - \dot{\delta}$ in the right leg. Transverse kinetic energy near the ends, and other end corrections to the kinetic energy, are neglected because $D \ll L$ is assumed. Then the total kinetic energy is

$$K = \frac{1}{2}(\rho_{\text{avg}} - \Delta\rho\delta/L)\frac{A}{2}L(-\omega a \sin\omega t + \dot{\delta})^2 + \frac{1}{2}(\rho_{\text{avg}} + \Delta\rho\delta/L)\frac{A}{2}L(-\omega a \sin\omega t - \dot{\delta})^2 \tag{22}$$

$$= \frac{AL}{2}\rho_{\text{avg}}(\omega^2 a^2 \sin^2\omega t + \dot{\delta}^2) + A\Delta\rho\delta\dot{\delta}\omega a \sin\omega t. \tag{23}$$

In Eq. (22), the first term is the kinetic energy in the left leg of the loop, and the second term that in the right leg. The density factors in these terms come from averaging Eq. (21) with respect to z.

The potential energy change U due to δ can be estimated by considering Fig. 2(b). As δ changes from zero to a nonzero value, isotherms far from the ends of the tube contribute no change to U, because for any mass moving up in the left leg there is an equal mass associated with the same isotherm moving down the same distance in the right leg. The same cancellation would occur for the gas within δ of the end of the tube, if it did not have to "turn the corner," changing from the left leg to the right leg at the top or the right to the left at the bottom; if such gas parcels could move to the positions shown as crosshatched in Fig. 2(b), their effects on U would be canceled by their partners of the same isotherms in the other leg. Thus, the net effect of nonzero δ is to lower some cold gas whose mass is of the order of $\rho_C(A/2)\delta$ a distance of the order of $D\sin\theta - \delta\cos\theta$ and raise some hot gas whose mass is of the order of $\rho_H(A/2)\delta$ a similar distance, yielding

$$U \simeq -\Delta\rho\frac{A\delta}{2}g\left(\frac{4}{3\pi}D\sin\theta - \delta\cos\theta\right), \tag{24}$$

where the $4/3\pi$ comes from careful consideration of the semicircular cross section of each leg.

With the standard Lagrangian methods of classical mechanics, the equation of motion for δ is obtained by writing $(d/dt)\partial[K-U]/\partial\dot{\delta} - \partial[K-U]/\partial\delta = 0$. Using Eqs. (23) and (24) above for K and U, the result is

$$\rho_{\text{avg}}AL\ddot{\delta} = \frac{A\Delta\rho g}{2}\left(\frac{4}{3\pi}D\sin\theta - 2\delta\cos\theta\right) - A\Delta\rho\omega^2 a\delta\cos\omega t. \tag{25}$$

This equation resembles Eq. (7) for the unbalanced-ring pendulum. The total mass $\rho_{\text{avg}}AL$ in the loop of piping accelerates in the δ direction in response to forces of gravity, expressed by the first term, and in response to forces caused by vibration, expressed by the second term. Following the same procedure as for the unbalanced-ring pendulum, this equation of motion is broken down into fast and slow parts by substituting $\delta = \delta_{\text{slow}} + \delta_{\text{fast}}\cos\omega t$ and assuming $\delta_{\text{fast}} \ll \delta_{\text{slow}}$ and $\omega^2 \gg \Delta\rho g/\rho_{\text{avg}}L$. The fast part of δ is then

$$\delta_{\text{fast}} = \frac{\Delta\rho}{\rho_{\text{avg}}}\frac{a}{L}\delta_{\text{slow}}, \tag{26}$$

and the slow response of δ to gravity and to the time-averaged product of δ_{fast} and the imposed vibration is described by

$$\rho_{\text{avg}} AL\ddot{\delta}_{\text{slow}} = \frac{A\Delta\rho g}{2}\left(\frac{4}{3\pi}D\sin\theta - 2\delta_{\text{slow}}\cos\theta\right) - \frac{A(\Delta\rho)^2\omega^2 a^2\delta_{\text{slow}}}{2\rho_{\text{avg}}L}. \tag{27}$$

If the vibrations suppress convection, then $\ddot{\delta}_{\text{slow}}=0$, and the phenomenon should be governed by the surviving terms on the right-hand side. Solving for δ_{slow} yields

$$\delta_{\text{slow}} = \frac{4D\sin\theta/3\pi}{\omega^2 a^2\Delta\rho/\rho_{\text{avg}}gL + 2\cos\theta}. \tag{28}$$

Too large a value of δ_{slow} would be unrealistic, because it would put the off-center cold gas and the off-center hot gas in Figs. 1(d) and 2(b) close together, thermally short-circuiting the temperature difference responsible for the vibration-stabilization effect. Thus, a stable δ_{slow} can be no larger than some fraction of L, which can be conveniently written as $2L/3\pi\alpha$, where α is as yet unknown. Making that substitution for δ_{slow} in Eq. (28), rearranging, and defining a dimensionless group of variables resembling Eq. (14) yield

$$N_{\text{ptc}} \equiv \frac{\omega^2 a^2}{g(\alpha D\sin\theta - L\cos\theta)}\frac{\Delta\rho}{\rho_{\text{avg}}} = 2. \tag{29}$$

Since $\Delta\rho/\rho_{\text{avg}}=\Delta T/T_{\text{avg}}$, this supports the dependences shown in Eq. (14) above, for $\beta=1$. Note that this derivation is valid for $90° \le \theta \le 180°$, so the geometrical factor in the denominator could just as well be written as $\alpha D|\sin\theta| + L|\cos\theta|$.

Equations (28) and (29) are only valid for ωa large enough to suppress convective motion. For very large ωa, δ_{slow} is generally small, as illustrated in Fig. 2(b). However, if ωa is just below the threshold, δ_{slow} could be fairly large and essentially time dependent, and the picture of Fig. 2(b) would be unrealistic because the off-center slugs of extreme-temperature gas would extend over appreciable lengths, and their temperatures would no longer be uniformly at T_C and T_H, but rather would be distributions of less-extreme temperatures determined by competing conduction to both heat exchangers and between the two legs of the loop. Whether this might lead to $N_{\text{ptc}} \sim (\Delta T/T_{\text{avg}})^{\beta}$, where $\beta<1$, is not clear. Further analysis of this issue may require numerical study of Eqs. (15)–(17) and other information in Ref. 11.

Repeating this section's analysis but starting with the assumption that $1/\rho \propto T$ is linear in z instead of the assumption of Eq. (21) that ρ is linear in z leads tediously to

$$\frac{(\Delta T)^2}{T_C T_H \ln(T_H/T_C)} \tag{30}$$

instead of $\Delta\rho/\rho_{\text{avg}}$ in Eq. (29). The difference between $\Delta T/T_{\text{avg}}$ and Eq. (30) is only $(\Delta T)^3/2T_{\text{avg}}^3$ to lowest order in $\Delta T/T_{\text{avg}}$. The accuracy of the measurements described below does not justify the extra complexity of Eq. (30), so we retain the simpler $\Delta T/T_{\text{avg}}$ and $\Delta\rho/\rho_{\text{avg}}$ dependences in Eqs. (14) and (29).

The high-amplitude stability of pulse tubes against gravity-driven convection was characterized by Wang and Gifford[8] in terms of the inverse of the dimensionless group

$$\frac{u_a^2}{gD\Delta T/T_{\text{avg}}} = \frac{\omega^2 a^2}{gD}\frac{T_{\text{avg}}}{\Delta T}, \tag{31}$$

which is similar to Eqs. (29) and (14), but with two important differences. First, the choice of Ref. 8 keeps g and ΔT together in the denominator, while our derivation of Eq. (29) shows that the nonlinear nature of the stabilizing effect of vibrations puts $(\Delta\rho)^2$ in the last term in Eq. (27) and, hence, puts $\Delta\rho$ in the numerator of Eq. (29), leaving g in the denominator: In contrast to the dependence shown in Eq. (31), higher temperature differences actually allow suppression of convection at *lower* frequencies and amplitudes, even while a larger acceleration of gravity would require higher amplitudes. Second, Ref. 8 arbitrarily chose D as the characteristic length in the dimensionless group, while our derivation shows that the characteristic length might best be considered to be θ dependent, and that L is more important than D when $D \ll L$, except very close to $\theta=90°$.

III. EXPERIMENTS

To investigate these phenomena under a broad range of experimental conditions, an apparatus with interchangeable tubes much simpler than complete pulse-tube refrigerators was built. Working only at and above ambient temperature allowed the use of easily measured electric-resistance heat, without refrigeration, and adoption of nearly standing-wave phasing for the measurements eliminated need for a pulse-tube refrigerator's orifice and compliance tank, simplified the apparatus, reduced surface areas that could contribute to room heat leaks, reduced the heat demands on the heat exchangers, and led to rapid thermal-equilibration times. Five tubes, shown to scale in Fig. 3 and described in Table I, were used for these measurements.

Each of the five pulse tubes (or thermal-buffer tubes) was a right-circular cylindrical space bounded around its sides by a 0.8-mm-thick stainless-steel wall and on its ends by diffusion-bonded stainless-steel screens acting as flow straighteners. Each flow straightener comprised 27 layers of nominally 16.5 wires/cm, 0.14-mm-diameter-wire square-weave screen, with alternate layers turned 45°. They were cut to a diameter that was 1.6 mm greater than that of each pulse-tube's inside diameter, by wire electric-discharge machining after diffusion bonding, so steps on the ends of the pulse tube could hold them firmly in place and define the pulse-tube length L accurately. Beyond these flow straighteners were drilled copper disks that served as heat exchangers, maintaining nearly isothermal planes across the ends of the flow straighteners by conducting heat to or from their surroundings. The heat-exchanger holes were 1.32 mm in diameter, and the hole patterns were designed for spatially uniform coverage over the pulse-tube area.

On the hot end, a bounce space the same diameter as the pulse tube allowed significant oscillating flow through the hot heat exchanger, and a 1.5-mm-diameter sheathed type-K thermocouple in that space, a few millimeters from the hot heat exchanger, measured T_H. The thermocouple was bent, as shown in the figure, so almost 1 cm of its tip lay close to the heat exchanger (except for the thinnest tube, in which the

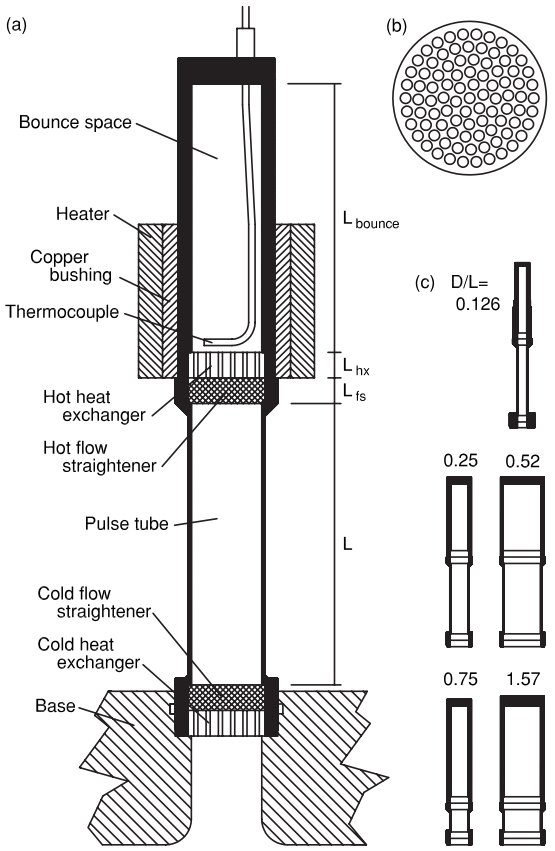


FIG. 3. (a) Cross-sectional scale drawing of the $D/L=0.25$ tube, shown with the cold end down (i.e., $\theta=0$). The dimensions L_j are given in Table I. The pressure-vessel boundary, shown in heavy black, was a long, machined tube with a cap welded into one end. The cap pressed on a thin-walled sleeve in the bounce space (not distinct in the figure), whose inside diameter was the same as that of the pulse tube, thereby trapping the hot heat exchanger and hot flow straightener against a machined step at the top of the pulse tube. The cap included a welded-in compression fitting through which the hot thermocouple passed. Clamps holding the tube to the base are not shown. (b) A perpendicular cross section though one of the drilled copper heat exchangers, at twice the scale of (a). (c) The proportions of all five tubes, at 1/4 the scale of (a).

bent portion was necessarily shorter). A commercial "band" electric-resistance heater provided heat, that heat being spread around the hot heat-exchanger region by a copper bushing. Ceramic-fiber thermal insulation covered all the hot parts including the pulse tube itself. The cold end was mounted in a water-cooled aluminum base, whose temperature T_C varied no more than 1 °C during the course of any single data set, and did not differ from 20 °C by more than a few degrees from week to week.

A passage through the aluminum base, a few centimeters long, led from the cold heat exchanger to the top of a 10-cm-diameter piston, which was driven by a linear motor[14] to whose housing the aluminum base was bolted. The motor was best operated very near the resonance frequency defined by its large moving mass and the gas-pressure spring constant experienced by the piston. This resonance frequency was easily varied by adjusting the mean pressure, and could be changed for a desired mean pressure by inserting volume-adding spacer rings between the motor housing and the aluminum base. The motor housing was mounted on a modified rotary stand, originally intended for rebuilding automobile engines. The rotary part of the stand had a hole-and-pin mechanism for reproducibly setting the tilt of the entire apparatus in 10.0° increments from 0.0° to 180.0°. A bubble level was used to align the pulse tube with gravity to $<0.1°$ with the apparatus set at 180.0°. The pressure amplitude p_a applied at the bottom of the pulse-tube assembly was measured with a lock-in amplifier connected to a piezoresistive transducer[15] in the aluminum base.

In the $D/L=0.52$ tube, a second thermocouple was installed, in the copper bushing under the electric-resistance heater. Near $T_H=250$ °C, the bushing thermocouple was never more than 10 °C hotter than the internal thermocouple, this temperature difference being largest when the convective heat transport was the largest.

TABLE I. Dimensions for the five tubes used in the experiments, and the heat $\dot{Q}_{\text{gascon}}$ carried by simple conduction in the gas in each tube under typical experimental circumstances. See also Fig. 3(c) for scale drawings.

$D/L=$	0.126	0.249	0.521	0.750	1.57
		Dimensions			
D (cm)	0.88	1.74	3.64	1.74	3.64
L (cm)	6.98	6.99	6.99	2.32	2.32
L_{fs} (cm)	0.64	0.64	0.64	0.64	0.64
φ_{fs}	0.82	0.82	0.82	0.82	0.82
L_{hx} (cm)	0.56	0.56	0.56	0.56	0.56
No. of hx holes	19	91	331	91	331
φ_{hx}	0.427	0.521	0.436	0.521	0.436
Holes' R_{hx} (mm)	0.66	0.66	0.66	0.66	0.66
L_{bounce} (cm)	6.66	6.74	6.74	9.02	9.02
		$\dot{Q}_{\text{gascon}}$ (W)			
He, $T_H=425$ °C		0.30			
He, $T_H=250$ °C	0.038	0.148	0.65	0.45	1.96
He, $T_H=150$ °C		0.076			
0.9He–0.1Ar, $T_H=250$ °C		0.121			
Ar, $T_H=250$ °C		0.019			

Obtaining one set of data typically took half of a day. A gas, its mean pressure p_m, a frequency, and a tip angle θ were chosen, and were kept fixed for each data set. An initial motor drive voltage was chosen, and heat was applied to the electric-resistance heater to maintain the hot temperature at a selected T_H. To assess that process, temperature was displayed as a function of time with a chart recorder. The heater voltage was adjusted manually until a steady setting achieved both a low rate of change in temperature (less than 0.1 °C in a few minutes) and the desired T_H. The steady-state heater voltage V and pressure amplitude p_a were then recorded, and the heater power $\dot{Q}$ was obtained by squaring the voltage and dividing the result by the heater's resistance. The motor drive voltage was changed to a new value, and the heat adjustment and data recording were repeated, typically at rates of two to four data points per hour.

Figure 4(a) shows six such data sets, all in the $D/L=0.25$ tube with 3.1-MPa helium gas driven at 100 Hz, and with $T_C=20$ °C and $T_H=250$ °C.

At $\theta=0°$, the cold end of the pulse tube was straight down so the gas was convectively stable. The measurements show that 14 W of heat were needed to keep $T_H=250$ °C in this tube, with amplitude-dependent variation being only a fraction of 1 W. Calculations show that the helium in the pulse tube conducted 0.15 W and the stainless-steel pulse-tube wall conducted 2.5 W, so most of the required heat was apparently heat leak through the fiber insulation to the room. Calculations[16] that include boundary-layer heat shuttle along the pulse tube and acoustic-power dissipation in the hot heat exchanger and flow straightener show that the required heat should drop quadratically by 0.8 W as the oscillation amplitude rises from $p_a/p_m=0.01$ to $p_a/p_m=0.05$, in rough agreement with the $\theta=0°$ measurements.

Compared with $\theta=0°$, only a little more heat was convected at $\theta=60°$ at zero or low oscillation amplitude. This tube was slender enough that the highest edge of its cold end was still 1.0 cm below the lowest edge of its hot end at $\theta=60°$, so the gas in the tube can still be regarded as convectively stable at this tip angle.

At $\theta=90°$, 120°, 150°, and 180° over 4 W of heat was convected through the tube when no oscillations were present, representing Nusselt numbers ranging from 30 at 90° to 50 at 150°. Such convection is large enough to reduce the cooling power of a pulse-tube refrigerator significantly. The Rayleigh number based on L is about 27×10^6, and such Nusselt numbers are plausible at this Rayleigh number: Eq. (4.89) in Ref. 17 yields a Nusselt number of 18 under these conditions, for $\theta=180°$. (However, our enclosure has porous ends, which could tend to increase the Nusselt number.) From the convective heat flow, ρ, c_p, and ΔT, we estimate that the convective velocity was of the order of 1 cm/s, roughly 100 times less than the typical oscillating velocity. The Reynolds number of the convective motion here is of the order of 20, so the convection should be laminar. This suggests that numerical calculations based on Ref. 11 may yield reliable results in this range of parameter space. However, in the tubes with $D/L\geq0.5$ we did sometimes see time-dependent convection, evidenced by time dependence in the hot temperature, whose variations were as high as a few tenths of a degree over time scales of about 10 s. The time dependence started near the convection-suppression transition and rose with amplitude, and was greatest for the tube with the highest D/L. Numerical calculations in the time-dependent regime might be more challenging.

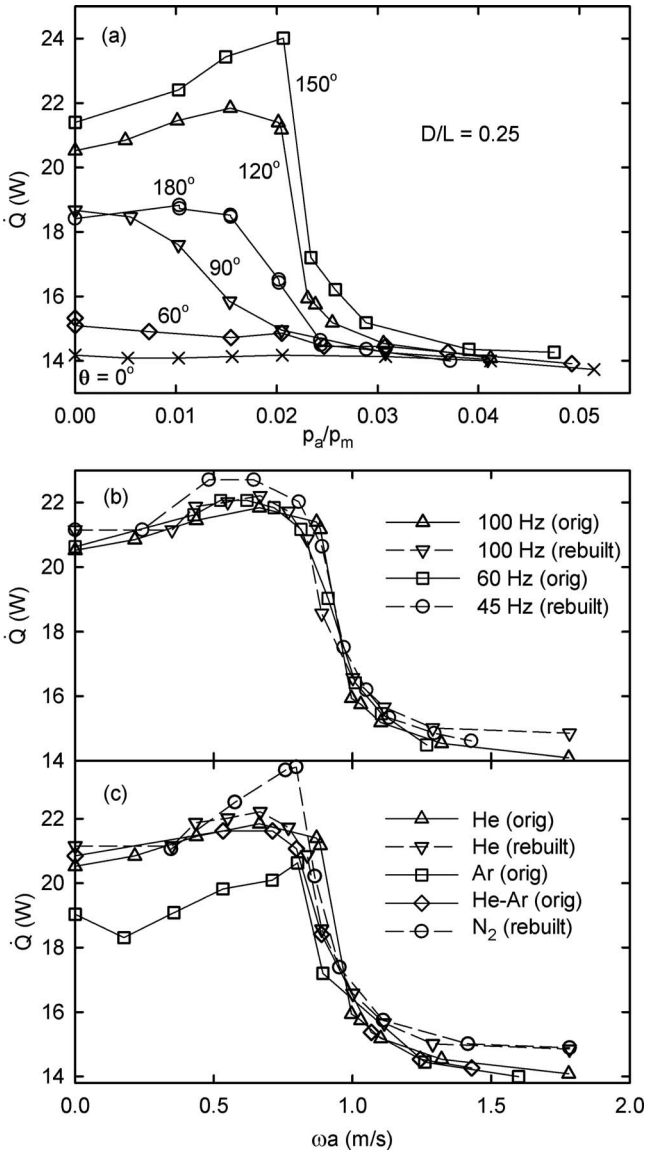


FIG. 4. Heat $\dot{Q}$ required to maintain a steady hot temperature under a wide variety of conditions with 3.1-MPa gas in the $D/L=0.25$ tube. The points are measurements, and the lines are only guides to the eyes. (a) Heat required to maintain $T_H=250$ °C, with helium at 100 Hz, for six tip angles θ. The horizontal axis is the pressure amplitude at the base, divided by mean pressure. (b) Heat required to maintain $T_H=250$ °C, with helium at $\theta=120°$, for three frequencies. (c) Heat required to maintain $T_H=250$ °C, at 100 Hz, at $\theta=120°$, for four different gases. The mixture was 90% helium and 10% argon. The horizontal axis in (b) and (c) is the square root of the relevant part of Eq. (14), because this keeps the points almost equally spaced, making the variations near the transition easier to see.

Figure 4(a) shows an initial rise in convective heat transfer with amplitude for $\theta\geq120°$. Possible explanations for this phenomenon include a weakening of the zero-velocity boundary condition at the ends of the convective cell as those ends find themselves, on average, farther from the flow straighteners at higher amplitude, and a strengthening of thermal contact near the ends of the convective cell as jets

whose diameters are of the order of the flow-straighteners' hydraulic radius squirt gas at the heat-exchangers' temperatures into the ends of the convective cell.

Figure 4(a) also shows that the convection was effectively stopped when the oscillations had a high enough amplitude, as expected from Secs. I and II.

The closely spaced points near the 120° transition and the essentially overlapping points throughout the 180° data indicate attempts to observe hysteresis, by taking some of the data while systematically increasing p_a and other data while systematically decreasing p_a. No hysteresis was observed in these data sets or in any others. [A near exception is described in the caption for Fig. 6(b).]

To plot such data as a function of ωa or of N_{ptc}, we converted from the measured p_a/p_m to the vibration amplitude a in the middle of the pulse tube by using

$$a = \frac{p_a}{\gamma p_m}\left(\frac{L}{2} + \gamma\varphi_{fs}L_{fs} + \left[1 + (\gamma - 1)\frac{\delta_\kappa}{R_{hx}}\right]\varphi_{hx}L_{hx} + L_{\mathrm{bounce}}\right), \tag{32}$$

where half of the pulse-tube length, $L/2$, the hot-flow-straightener length L_{fs}, the hot-heat-exchanger length L_{hx}, and the bounce-space length L_{bounce} add up to the total distance between the middle of the pulse tube and the closed end of the experiment. This expression is based on the assumptions that p_a is independent of x from the middle of the pulse tube to the end of the bounce space, and that thermal-hysteresis effects in the bounce space and pulse tube are negligible. Thus, if the total distance were unobstructed and of uniform cross-sectional area, then Eq. (32) would be simply $a=(p_a/\gamma p_m)(L/2+L_{fs}+L_{hx}+L_{\mathrm{bounce}})$, describing simple adiabatic compressions and expansions everywhere. The prefactor $p_a/\gamma p_m$ is used in Eq. (32) because most of that length, $L/2+L_{\mathrm{bounce}}$, does experience nearly adiabatic compressions and expansions. The prefactors of the small L_{fs} and L_{hx} terms in Eq. (32) account for the volumetric porosities φ_j of those components, the isothermal nature of the oscillations in the flow straightener, and the thermal hysteresis in the circular channels through the heat exchanger, in which the channel radius is R_{hx} and the thermal penetration depth is δ_κ. Numerical estimates[16] that include inertial and resistive pressure drops in the hot heat exchanger and flow straightener and the consequences of thermal hysteresis elsewhere suggest that these assumptions introduce errors of no more than 2% to the determination of a.

Figure 4(b) shows convection-suppression data from the $D/L=0.25$ tube at three different frequencies, all with 3.1-MPa helium at $\theta=120°$ and $T_H=250$ °C. Although the frequency ranges over a factor of 2, plotting these data sets as functions of ωa aligns them very well, corroborating the ωa functional dependence on N_{ptc} in Eq. (14), and contradicting any other supposed strong dependences on ω or a in this tube, such as a/L (independent of ω) or $\omega a^2/L$.

[The "100 Hz (orig)" data set in Fig. 4(b) is also shown in Fig. 4(a). After taking that data set and the "60 Hz (orig)" set, the original $D/L=0.25$ tube was disassembled to use parts elsewhere, and later was "rebuilt" to obtain more data. The difference between the "original" and "rebuilt" 100-Hz sets is presumably due to slight hardware irreproducibility, including slightly different hot-thermocouple positions. For future work, we recommend a reproducible attachment of both thermocouples directly to their copper heat exchangers.]

Figure 4(c) shows convection-suppression data from the $D/L=0.25$ tube for four different gases at 3.1 MPa and 100 Hz, with $T_H=250$ °C and $\theta=120°$. The horizontal alignment of all of these data sets confirms the lack of explicit gas-property dependence of N_{ptc}. The alignment of the helium and argon data, despite the tenfold difference in atomic mass and mass density, confirms that N_{ptc} should be independent of molecular mass. The alignment of the helium-argon data with the pure-monatomic-gas data, despite the mixture's 32% lower Prandtl number, confirms that N_{ptc} is independent of Prandtl number and, by inference, independent of the gas transport properties. The alignment of the $\gamma=7/5$ nitrogen data with the $\gamma=5/3$ monatomic-gas data confirms that N_{ptc} is independent of the specific-heat ratio, except through the conversion from p_a/p_m to a given in Eq. (32).

To investigate the ΔT dependence of the convection-suppression transition, we used the $D/L=0.25$ tube with 3.1-MPa helium and $\theta=120°$ at 100 Hz, at three different hot temperatures. To bring the data into approximate *vertical* alignment, we divided $\dot{Q}$ by ΔT, and then subtracted 0.08, 0.06, and 0.05 W/°C from the 425, 250, and 150 °C data, respectively, to account for the temperature-dependent heat leaks. With the three data sets plotted against ωa in Fig. 5(a), it is apparent that it was easier to suppress the convection at higher ΔT. Figures 5(b) and 5(c) show these three data sets plotted against $\omega a\sqrt[4]{\Delta T/T_{\mathrm{avg}}}$ and $\omega a\sqrt{\Delta T/T_{\mathrm{avg}}}$. The data align best using the fourth root, which is why we choose $\beta=1/2$ in Eq. (14), despite the fact that the derivation of Sec. II yields $\beta=1$.

To study the L and D dependence of the convection-suppression transition, we used data from all five pulse tubes, which had five different aspect ratios. Measurements with identical gas, temperatures, and frequency are shown in Figs. 4(a) and 6. Like the $D/L=0.25$ tube, which yielded the data shown in Figs. 4 and 5, the $D/L=0.52$ tube displayed sharp convection-suppression transitions at $\theta=120°$ and 150°, and a $\theta=0$ heat requirement that was almost independent of amplitude, as shown in Fig. 6(b). In the $D/L=0.75$ tube, the transitions were still sharp, but the $\theta=0$ heat requirement rose dramatically, and quadratically, with amplitude, as shown in Fig. 6(c). The $D/L=1.57$ tube showed a similar rising baseline heat requirement, but a very ill-defined and incomplete transition to reduced convection, as shown in Fig. 6(d). Our motor did not let us learn whether higher amplitudes would bring a second, more complete reduction in convection in this tube. Unlike the other four tubes, the $D/L=0.126$ tube did not have sharp transitions at any tip angle, as shown in Fig. 6(a), and the small heats involved were difficult to measure accurately.

We do not understand some of these qualitative differences between the data sets in the different tubes. The quadratically rising $\theta=0$ heats for the two shortest tubes suggest streaming, but the calculated Rayleigh streaming velocity[18] just outside the boundary layer at mid-tube is very nearly the same, 1.3 cm/s at $p_a/p_m=0.025$, for all five tubes, and esti-

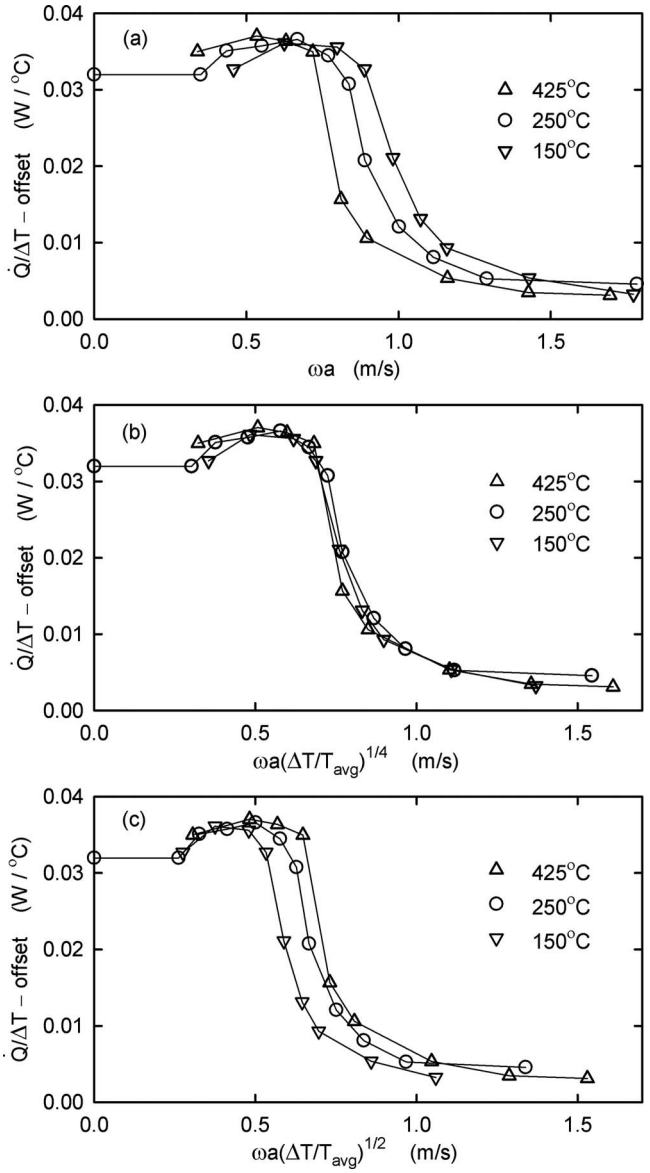


FIG. 5. Normalized heat required to maintain a steady hot temperature in the (rebuilt) $D/L=0.25$ tube, with 3.1-MPa helium and $\theta=120°$, at three different hot temperatures T_H with which the points are labeled. (N.B.: ΔT is ~20 °C smaller than T_H.) The points are measurements, and the lines are only guides to the eyes. The experimental temperature dependence is closer to the fourth root used in (b) than to either the square root used in (c) or no temperature dependence used in (a).

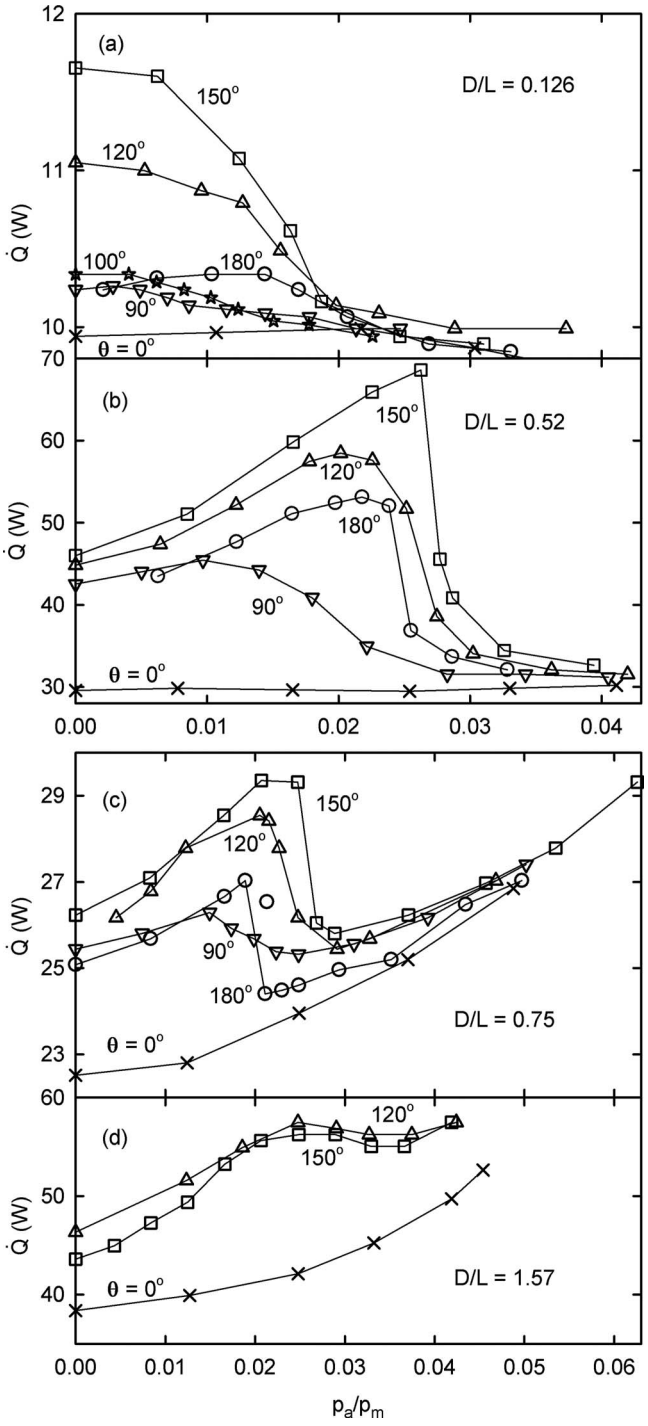


FIG. 6. Heat required to maintain $T_H=250$ °C in 3.1-MPa helium at 100 Hz, in four different tubes. Figure 4(a) shows similar data for a fifth tube. The points are measurements, and the lines are only guides to the eyes. The horizontal axis is pressure amplitude at the base, divided by mean pressure. In (c), one of the 180° points is not connected with the lines. That data point was metastable: It persisted steadily for 5 min before the heat suddenly dropped to the point below it.

mates of the heat that such streaming can transport along the tubes[19] range only from 0.1 to 0.5 W at that p_a/p_m, too small to explain the measurements. Seeking another reason that the short tubes differ from the long tubes, one can consider the stroke divided by the tube length, $2a/L$, which should be smaller than 1 to prevent gas from shuttling heat all the way from the hot flow straightener to the cold flow straightener every cycle of the oscillation. But $2a/L$ is only 0.12 at the $\theta=120°$ transition in the $D/L=0.75$ tube, where the rising baseline is perhaps even visible as low as $2a/L\simeq0.07$, so shuttle heat should not be responsible for the rising baseline. Furthermore, the $D/L=0.25$ tube's 100-Hz, $\theta=0$ data reach as high as $2a/L=0.09$, and that tube's 45-Hz data extend to $2a/L=0.14$, with no suggestion of rising baselines in Figs. 4(a) or 4(b).

Despite this mystery, the data from the four tubes with the smallest D/L can be used to explore whether D or L is the most important geometrical variable l in Eq. (20) governing the convection-suppression transition, and to test the $\alpha D \sin\theta - L\cos\theta$ geometry dependence given in Eq. (29) for small D/L. From each $\theta\neq0$ data set in Figs. 4(a) and 5, we subtracted a quadratic fit to the corresponding $\theta=0$ base-

line data set, and defined the transition from convection to suppression as the value of p_a/p_m where each data set passes halfway between the maximum value of $\dot{Q}-\dot{Q}_{\text{baseline}}$ and zero. (This definition of the transition is essentially identical to the location of steepest decrease in $\dot{Q}-\dot{Q}_{\text{baseline}}$, except for the $D/L=1.57$ tube.) We converted the transition value of p_a/p_m to a corresponding transition value of a using Eq. (32).

Figure 7 displays the results as a function of D/L for three different choices of the characteristic length l that might be used in the dimensionless group in Eq. (20). First, Fig. 7(a) shows the results when plotted with $l=D$, the choice made in Ref. 8. For this choice of l, the transition displays complicated dependence on θ and D/L, suggesting that simply using $l=D$ in N_{ptc} does not provide a universal description of the transition. In fact, for $\theta=180°$, the transition varies almost as $1/(D/L)$, as shown by the dashed curve, suggesting that dividing by D is a very poor choice for this particular θ. Next, in Fig. 7(b), the same data are plotted using $l=L$. Here, the $\theta\neq 90°$ data collapse reasonably well along a single curve with little D/L dependence, but the $\theta=90°$ data deviate significantly from the others; comparison to the dashed line shows that the $\theta=90°$ transition varies almost as D/L for small D/L, suggesting that dividing by L is a poor choice for this particular θ. Finally, Fig. 7(c) shows the same data plotted using $l=\alpha D\sin\theta-L\cos\theta$, with $\alpha=1.5$. This choice brings the data sets for all tip angles close to a common curve, consistent with Eq. (29) in some ways. Trying $\alpha=1.0$, 2.5, or more-extreme values ruins the clustering of the data in Fig. 7(c), while using $\alpha=2.0$ looks only a little worse than $\alpha=1.5$. Using $\alpha=1.5$ sets $\delta_{\text{slow}}=2L/3\pi\alpha=0.14L$, which seems reasonable, being about three times larger than the δ_{slow} shown in Fig. 2.

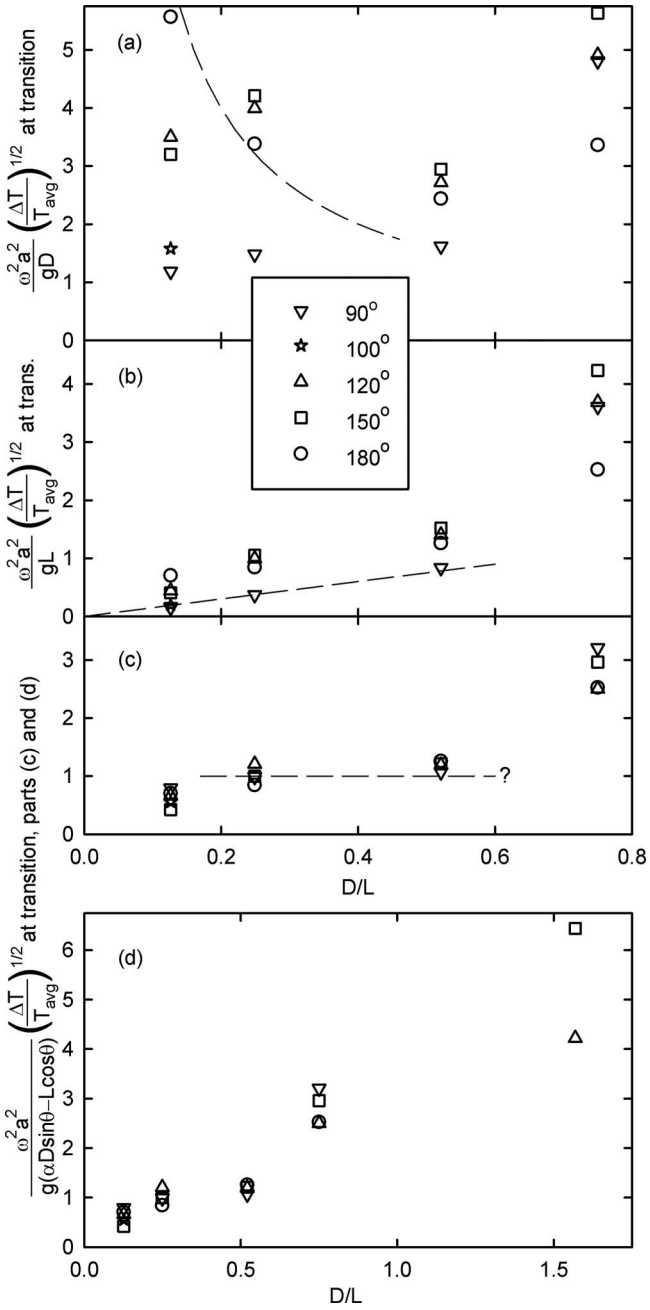


FIG. 7. Transitional values of N_{ptc} based on data from Figs. 4(a) and 6, testing different choices for l in Eq. (20). (a) Choosing $l=D$ yields transitional values of N_{ptc} that are not independent of aspect ratio at small aspect ratio when $\theta\neq 90°$. The dashed curve shows $N_{\text{ptc}}\propto 1/(D/L)$, for comparison with the $\theta=180°$ data. (b) Choosing $l=L$ yields transitional values of N_{ptc} that are more nearly independent of aspect ratio at small aspect ratio when $\theta\neq 90°$, but leaves the $\theta=90°$ data varying almost linearly with aspect ratio, as suggested by the dashed line. (c) Choosing $l=\alpha D\sin\theta-L\cos\theta$, with $\alpha=1.5$, yields transitional values of N_{ptc} that are more nearly independent of both θ and D/L. (d) For completeness, this shows the same data and vertical axis as in (c), but with the data from the highest-aspect-ratio tube included.

IV. FURTHER DISCUSSION

The vibrational stabilization of an inverted pendulum is a useful guide to intuition about how acoustic oscillations suppress natural convection in an inverted pulse tube, and the dimensionless pulse-tube convection number N_{ptc} defined in Eq. (14) may provide a good quantitative framework for analysis, at least for small aspect ratios. Experiments confirm that ωa captures the relevant dependences on frequency and displacement, and that gas properties such as γ and Prandtl number are not important. However, the picture is incomplete, at best.

For example, the observed $\beta=1/2$ temperature dependence in Fig. 5 differs significantly from the $\beta=1$ prediction of Eq. (29). This remains a mystery. In the same figure, dividing $\dot{Q}$ by ΔT brought the data into good *vertical* alignment, implying that the Nusselt number is independent of ΔT, while Eq. (4.89) in Ref. 17 predicts that the Nusselt number should be proportional to $(\Delta T)^{1/3}$.

Furthermore, we are not sure how to interpret the D/L dependence that remains in Figs. 7(c) and 7(d). One possibility is that the transition occurs at $N_{\text{ptc}}\simeq 1$ for a substantial range of D/L, including $0.25\leq D/L\leq 0.52$, as predicted for low D/L by Eq. (14) and suggested by the dashed line in Fig. 7(c). The data at $D/L=0.126$ might fall below this value because of physics not included in Sec. II: For example, in the $D/L=0.126$ tube, transverse thermal relaxation is faster than in any other tubes, and is comparable to the $\sqrt{l/g}$ convective-motion time, both because the helium-column diameter is smaller and because of the relatively greater contribution of circumferential conduction by the stainless-steel tube wall. A second interpretation would simply discount the

$D/L=0.126$ data because the convection there was weak, differed qualitatively from the other data sets (having no initial rise of $\dot{Q}$ with p_a), and was hard to measure well (e.g., day-to-day and hour-to-hour variations in room temperature would have had a greater effect on this data set than on the others). A third interpretation would be that the analysis of Sec. II is wrong and the data show that $N_{\mathrm{ptc}} \sim c_1 + c_2 D/L$ describes the transition for all D/L, with the data at $D/L = 0.52$ to be discounted for some unknown reason.

Resolving these and the other interesting, unanswered questions raised here may require additional experiments, numerical modeling, or both. One important question is whether the suppression of convection depends on the oscillating pressure at all. The time phase difference between the oscillating pressure and oscillating velocity in these experiments was 90°, while practical pulse tubes, transmitting acoustic power, operate closer to a time phase of 0° or 180°, in some cases with the time phase tuned to reduce Rayleigh streaming.[18] Whether this time phase affects the convection suppression, either directly or indirectly through Rayleigh streaming, has not been investigated here. And the magnitude of the oscillating pressure, neglected here in the discussion between Eqs. (20) and (21), might have a significant effect via the gas compressibility, because it makes the oscillating velocity at the ends of the pulse tube different from that in the center.

This situation is most unclear for $D/L > 1$, where Fig. 6(d) shows that the suppression of convection by vibration is very incomplete, or, at best, only partially completed at the amplitudes accessible in this experiment. At such aspect ratios, and with a sometimes a significant fraction of L, ensuring that imperfect flow straightening at the ends of the tube does not affect the measurements may be particularly challenging.

Other well-known rigid-pendulum phenomena, such as parametric resonance and synchronized unidirectional rotation,[1] may also have analogs in pulse tubes, at lower frequencies than those studied here.

ACKNOWLEDGMENTS

This work was supported by license income from thermoacoustic patents at Los Alamos National Laboratory. We thank Robert Keolian and Robert Ecke for many insights about the effects of vibrations on pendula and fluids, and for introductions to the relevant literature.

[1] A recent summary and introduction to some of the relevant literature are given by E. I. Butikov, "On the dynamic stabilization of an inverted pendulum," Am. J. Phys. **69**, 755–768 (2001).
[2] L. Blitzer, "Inverted pendulum," Am. J. Phys. **33**, 1076–1078 (1965).
[3] L. D. Landau and E. M. Lifshitz, *Mechanics* (Pergamon, New York, 1960), Sec. 30.
[4] D. Morin, *Introduction to Classical Mechanics* (Cambridge University Press, Cambridge, 2008), Solution 6.5.
[5] R. Radebaugh, "Development of the pulse tube refrigerator as an efficient and reliable cryocooler," Proceedings of the Institute of Refrigeration, London (2000), pp. 11–29; http://cryogenics.nist.gov/Papers/Institute_of_Refrig.pdf (Last viewed 9/30/2009).
[6] G. Thummes, M. Schreiber, R. Landgraf, and C. Heiden, "Convective heat losses in pulse tube coolers: Effect of pulse tube inclination," in *Cryocoolers 9*, edited by R. G. Ross (Plenum, New York, 1997), pp. 393–402.
[7] G. Thummes and L. Yang, "Development of Stirling-type pulse tube coolers driven by commercial linear compressors," in *Infrared Technology and Applications XXVIII*, edited by B. F. Andresen, G. F. Fulop, and M. Strojnik (Society of Photo-Optical Instrumentation Engineering, Seattle, WA, 2003), pp. 1–14.
[8] C. Wang and P. E. Gifford, "A single-stage pulse tube cryocooler for horizontally cooling HTS MRI probe," in *Advances in Cryogenic Engineering: Transactions of the Cryogenic Engineering Conference—CEC*, edited by J. Waynert, J. Barclay, S. Breon, E. Daly, J. Demko, M. DiPirro, J. R. Hull, P. Kelley, P. Kittel, A. Klebaner, J. Lock, J. Maddocks, J. Pfotenhauer, C. Rey, and Q.-S. Shu (American Institute of Physics, New York, 2004), Vol. **49**, pp. 1805–1811.
[9] S. Backhaus and G. W. Swift, "A thermoacoustic-Stirling heat engine: Detailed study," J. Acoust. Soc. Am. **107**, 3148–3166 (2000).
[10] D. L. Gardner and G. W. Swift, "A cascade thermoacoustic engine," J. Acoust. Soc. Am. **114**, 1905–1919 (2003).
[11] G. Z. Gershuni and D. V. Lyubimov, *Thermal Vibration Convection* (Wiley, New York, 1998).
[12] G. W. Swift, *Thermoacoustics: A Unifying Perspective for Some Engines and Refrigerators* (Acoustical Society of America, Sewickley, PA, 2002).
[13] G. W. Swift, *Thermoacoustics: A Unifying Perspective for Some Engines and Refrigerators* (Ref. 12), Eq. (7.53).
[14] Model C2, Clever Fellows Innovation Consortium, Inc., Troy, NY, http://qdrive.com (Last viewed 9/30/2009).
[15] Model 8510B-500M37, Endevco, San Juan Capistrano, CA, www.endevco.com (Last viewed 9/30/2009).
[16] B. Ward, J. Clark, and G. Swift, "Design environment for low-amplitude thermoacoustic energy conversion (DeltaEC)," software and user's guide available from the Los Alamos thermoacoustics web site http://www.lanl.gov/thermoacoustics/ (Last viewed 9/30/2009).
[17] W. M. Rohsenow, J. P. Hartnett, and Y. I. Cho, *Handbook of Heat Transfer* (McGraw-Hill, New York, 1998).
[18] J. R. Olson and G. W. Swift, "Acoustic streaming in pulse tube refrigerators: Tapered pulse tubes," Cryogenics **37**, 769–776 (1997).
[19] K. I. Matveev, S. Backhaus, and G. W. Swift, "The effect of gravity on heat transfer by Rayleigh streaming in pulse tubes and thermal buffer tubes," in *Proceedings of the IMECE 2004: International Mechanical Engineering Congress and Expo*, Fairfield, NJ (American Society of Mechanical Engineers (ASME), New York, 2004), pp. 125–139, Paper No. IMECE 2004-59076.

Broadband cluster transducer for underwater acoustics applications[a)]

Richard A. G. Fleming,[b)] Dennis F. Jones, and Charles G. Reithmeier
Defence Research and Development Canada—Atlantic, P.O. Box 1012, Dartmouth, Nova Scotia B2Y 3Z7, Canada

(Received 16 March 2009; revised 5 July 2009; accepted 4 August 2009)

A broadband cluster transducer, based on barrel-stave flextensional transducer technology, was designed, built, and tested in air and seawater at Defence R&D Canada—Atlantic. The design goal was to develop a transducer that exhibited a transmitting voltage response of greater than 120 dB//1 μPa-m/V from 1 to 7 kHz and have the potential for beamsteering. Six orthogonal piezoceramic-driven class I barrel-stave transducer elements mounted on a central manifold were wired separately, allowing each to be driven individually or in combination with the other elements. The resonance modes under the following drive conditions were determined from in-air conductance measurements: each of the six elements excited individually as mass-loaded class I flextensionals, three collinear pairs driven separately as class III flextensionals, and all six elements driven in phase simultaneously. The fully-driven cluster was found to have a transmitting voltage response of greater than 120 dB//1 μPa-m/V from 800 Hz to more than 10 kHz and its topology is amenable to beamsteering that is yet to be characterized. [DOI: 10.1121/1.3216911]

PACS number(s): 43.38.Fx [AJZ] Pages: 2285–2293

I. INTRODUCTION

Work is underway at Defence R&D Canada—Atlantic (DRDC Atlantic) to develop high-power and broadband transducers for applications including underwater communications, broadband mine detection, and antisubmarine warfare.[1,2] Recently, a broadband cluster transducer based on the symmetric dual-shell barrel-stave transducer technology was constructed at DRDC Atlantic for applications requiring both broadband and directional capabilities. Over the past 2 decades, several versions of the barrel-stave flextensional transducer were designed, built, and tested at DRDC Atlantic to support a variety of naval research and development activities in underwater acoustics.[3–6] Compact, lightweight, and cylindrical in form, these low-frequency sound sources were well suited to deployable naval sonar systems and subsequently patented both in Canada and the United States.[7] In 1989, a finite element analysis model was developed for the first DRDC Atlantic broadband dual-shell barrel-stave flextensional transducer. This transducer was an asymmetric design in that two radiating shells of different lengths were mechanically coupled through a common center plate. After construction, this transducer was mounted inside an oil-filled hose, and then tested in seawater at the DRDC Atlantic Acoustic Calibration Barge and in a submersible fiberglass pressure vessel at the NAVSEA Dodge Pond Acoustic Measurement Facility in Niantic, CT. Some of the performance results from the free-field tests at DRDC Atlantic were published in Ref. 4 and demonstrated that this asymmetric dual-shell transducer could survive water depths up to 280 m without a pressure compensation system like the one described by Bonin and Hutton.[8]

Later, a symmetric version (identical radiating shells) of the dual-shell barrel-stave transducer was tested and published by Jones and Reithmeier.[9]

Given the expertise in barrel-stave transducer development at DRDC Atlantic, a project was initiated to develop a broadband transducer that exhibited a transmitting voltage response (TVR) of greater than 120 dB//1 μPa-m/V from 1 to 7 kHz with potential for beamsteering and based on the barrel-stave design. In this paper, the broadband performance of the cluster transducer is presented using electroacoustic measurements performed in both air and seawater. This transducer's directivity patterns as a function of both frequency and active element combinations are to be carried out with both in-water measurements and finite element analysis. The cluster transducer shows promise in numerous naval applications including broadband communications, multistatic antisubmarine warfare, and broadband low-frequency mine detection.

II. BARREL-STAVE FLEXTENSIONAL TRANSDUCER CLASSIFICATION

Barrel-stave flextensional transducers are concave versions of the first three classes in the Brigham–Royster flextensional classification scheme,[10,11] as described in detail by Jones *et al.*[12] and Jones and Lindberg.[13] Cross-sectional sketches of three classes of DRDC Atlantic barrel-stave transducers are shown in Fig. 1. Since the transducers shown have several identical components, only seven labels are used to avoid clutter (i.e., a–g).

The class I barrel-stave transducer at the top of Fig. 3 has an internal piezoelectric ring-stack driver (a) made from

[a)]Portions of this work were presented in "Broadband response of a flextensional cluster transducer," at the 2008 U.S. Navy Workshop on Acoustic Transduction Materials and Devices, State College, PA, 13-15 May 2008.

[b)]Author to whom correspondence should be addressed. Electronic mail: richard.fleming@drdc-rddc.gc.ca

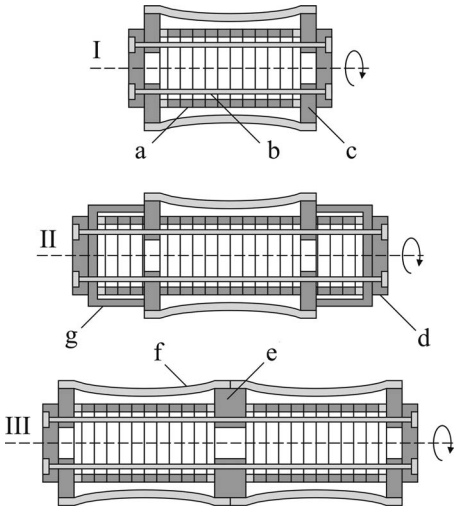


FIG. 1. Typical cross sections of class I, class II, and class III barrel-stave flextensional transducers with the rotational symmetry axes shown by the dashed lines and curved arrows. The class I design is driven by a piezoceramic ring stack (a) which is mechanically biased using stainless steel stress bolts (b). The bolts pass through mild steel end plates (c) and are fastened to aluminum end caps (d). The class II design has an extended ring-stack driver that is enclosed within mild steel housings (g). The class III design contains two ring-stack drivers connected to a mild steel center plate (e). Class I and class II designs have a set of concave aluminum staves (f) while the class III design has two sets.

axially-poled lead zirconate titanate ceramic rings.[14] Mechanical bias is applied to the ceramic ring stack using stainless steel bolts (b) to prevent the driver from going into tension during operation.[15,16] The stress bolts pass through central holes in mild steel end plates (c) located at the ends of the ring stack, and are fastened to aluminum end caps (d). The class II transducer in the center of Fig. 1 has a longer ring-stack driver that extends beyond the end plates and is contained inside two stiff mild steel housings (g). The greater volume of ceramic in the class II design distinguishes this class as a high-power version of the class I design. Of course, other high energy-density materials can be used in place of piezoceramic to produce greater acoustic source levels without going to a class II design. An example of this for a class I barrel-stave flextensional transducer is the Terfenol-D magnetostrictive alloy driver developed by Moffett and Clay.[17] The class III transducer at the bottom of the figure is a graft of two class I transducers sharing a common mild steel center plate (e). The flexural resonance modes of these transducers arise from the radial displacements of the concave set of aluminum staves (f) that surround the driver. The class III transducer has two sets of staves, which give rise to a broadband response.

The primary radiating surface of the class I barrel-stave transducer is comprised of eight concave aluminum staves, which are attached to octagonal mild steel end plates. A single stave is shown in the middle of the Fig. 2. On the right is an aluminum slotted shell, an alternative to staves and described in more detail by Merchant.[18] A photograph of two class III barrel-stave flextensional transducers is shown in Fig. 3. The dual-shell design, consisting of 16 staves, can be seen on the left. The transducer on the right is sealed with a neoprene rubber boot to prevent seawater ingress through the

FIG. 2. The DRDC Atlantic class I barrel-stave flextensional transducer on the left has eight concave aluminum staves that are attached to an octagonal mild steel end plate. The slotted-shell with a circular cross-section (left) can also be used as the radiating shell (see Ref. 18). Photograph by H. Merklinger, DRDC Atlantic, 2002.

gaps between adjacent staves. Since the class II barrel-stave flextensional transducer is not relevant to the broadband behavior of the flextensional cluster transducer, it is not within the scope of this paper. The interested reader can find design and performance information on the class II transducer in Refs. 12 and 13.

III. BROADBAND CLUSTER TRANSDUCER CONSTRUCTION

At the center of the cluster transducer shown in Fig. 4 is a manifold (e) fabricated from a solid mild steel cube. Machined into each of the six manifold faces is an octagonal flange used to support six class I barrel-stave elements labeled E1–E6. Holes through the manifold permit the passage of six pairs of electrical leads that exit the cluster at the end of element E1 and are used to drive one or more of the barrel-stave transducer elements.

The piezoelectric driver (a) for each of the barrel-stave cluster elements consists of a stack of ten Navy Type III lead zirconate titanate ceramic rings connected electrically in parallel and insulated at each end by machinable glass-ceramic rings. The outside surface of each stack is fiberglass wrapped to provide further electrical insulation. Each driver is mechanically biased using stainless steel stress bolts (b). The outside ends of the six piezoceramic drivers are bonded to mild steel end plates (c). An aluminum end cap (d) is bonded to the outside of each end plate to support the stress bolts. This can be seen at the end of element E2 in the photograph

FIG. 3. Two DRDC Atlantic Class III dual-shell barrel-stave flextensional transducers. Note that the transducer on the right is sealed with a neoprene rubber boot, which is required on all barrel-stave designs to prevent seawater ingress. Photograph by Donald Glencross, DRDC Atlantic, 2004.

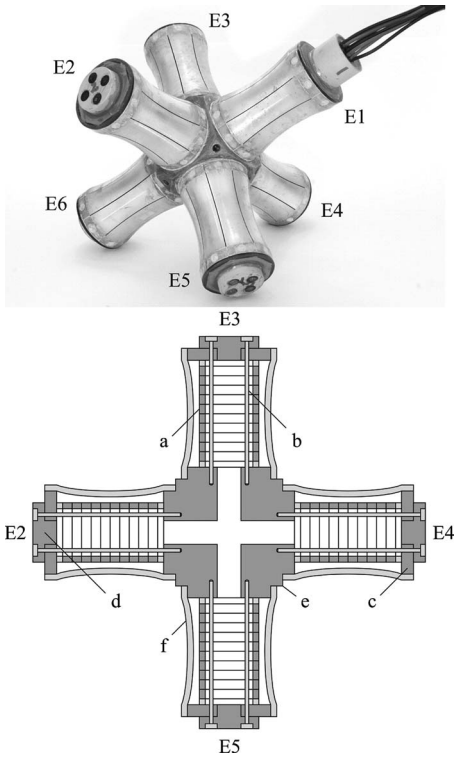


FIG. 4. The six-element flextensional cluster transducer without a neoprene rubber boot is shown at the top. Note that six sets of electrical leads exit through element E1. Shown at the bottom is a cross section in the E2-E5 plane. The component labels are the same as those in Fig. 1 with the exception that the mild steel center plate (e) in the class III cross section is replaced with a mild steel manifold (e) in the cluster transducer. Photograph by Donald Glencross, DRDC Atlantic, 2008.

in Fig. 4. The end cap for E1 is longer than the other five in order to support the electrical leads exiting the cluster transducer.

Bonded to the faces of the octagonal end plates and manifold flanges are sets of eight concave aluminum staves (f), with 48 staves in total. A gap is left between adjacent staves (see the photograph in Fig. 4). In order to prevent seawater ingress through these gaps, the transducer elements are sealed with neoprene rubber boots and all exposed surfaces are coated with polyurethane, as shown in Fig. 5. The mass and end-to-end dimensions of the fully assembled cluster transducer are 17.5 kg and 37 cm, respectively.

IV. FLEXURAL AND LONGITUDINAL MODES IN AIR

The in-air electrical conductance (G) of class I and class III barrel-stave flextensional transducers and the flextensional cluster transducer was measured with an Agilent 4294A precision impedance analyzer over the 40 Hz–10 kHz frequency band. These measurements are shown in Figs. 6–9. The frequencies and conductance values at resonance for the flexural and longitudinal modes are listed in Table I. For class I and class III barrel-stave flextensional transducers, the conductance was measured prior to the installation of the neoprene rubber boots, while the measurements on the cluster transducer were done after the boot was installed. In the case of the cluster transducer, ten combinations of barrel-stave elements were driven: (a) E1–E6 driven separately as mass-loaded class I transducers where the mass loading is provided by the manifold and five undriven elements; (b) E1 and E6, E2 and E4, and E3 and E5 driven as class III collinear pairs; and (c) all six elements E1–E6 driven simultaneously.

FIG. 5. The flextensional cluster transducer protected by a neoprene rubber boot. Photograph by Donald Glencross, DRDC Atlantic, 2008.

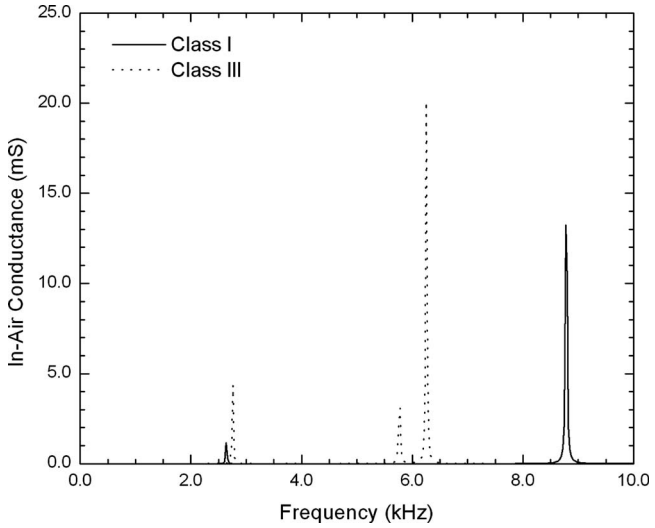


FIG. 6. In-air electrical conductance of the class I and class III barrel-stave flextensional transducers.

A. Class I barrel-stave flextensional transducer

The class I barrel-stave flextensional transducer shown at the top of Fig. 1 and on the left in Fig. 2 has a fundamental flexural mode whose frequency f_0 is primarily determined by the geometrical and material properties of the concave aluminum staves.[19] In addition, the axial displacements of the internal ring-stack driver give rise to a longitudinal mode whose frequency f_S varies inversely with the length of the driver. The subscript S represents the case where both ends of the driver, including the end plates and caps, are symmetric, or nearly so. Both resonance peaks appear in the conductance plot in Fig. 6 as the solid line. The fundamental flexural and symmetric longitudinal frequencies f_0 and f_S are 2.63 and 8.77 kHz, respectively (see Table I). The associated electrical conductance values G_0 and G_S are 1.16 and 13.2 mS, respectively. The physics of the various modes of the barrel-stave transducer are well described using equivalent circuit methods and finite element analysis by Moffat *et al.*[20] ATILA™ finite element code-generated[21] two-dimensional (2D) in-water fundamental flexural mode and first longitudinal mode representations can be seen in Figs. 10 and 11, respectively.

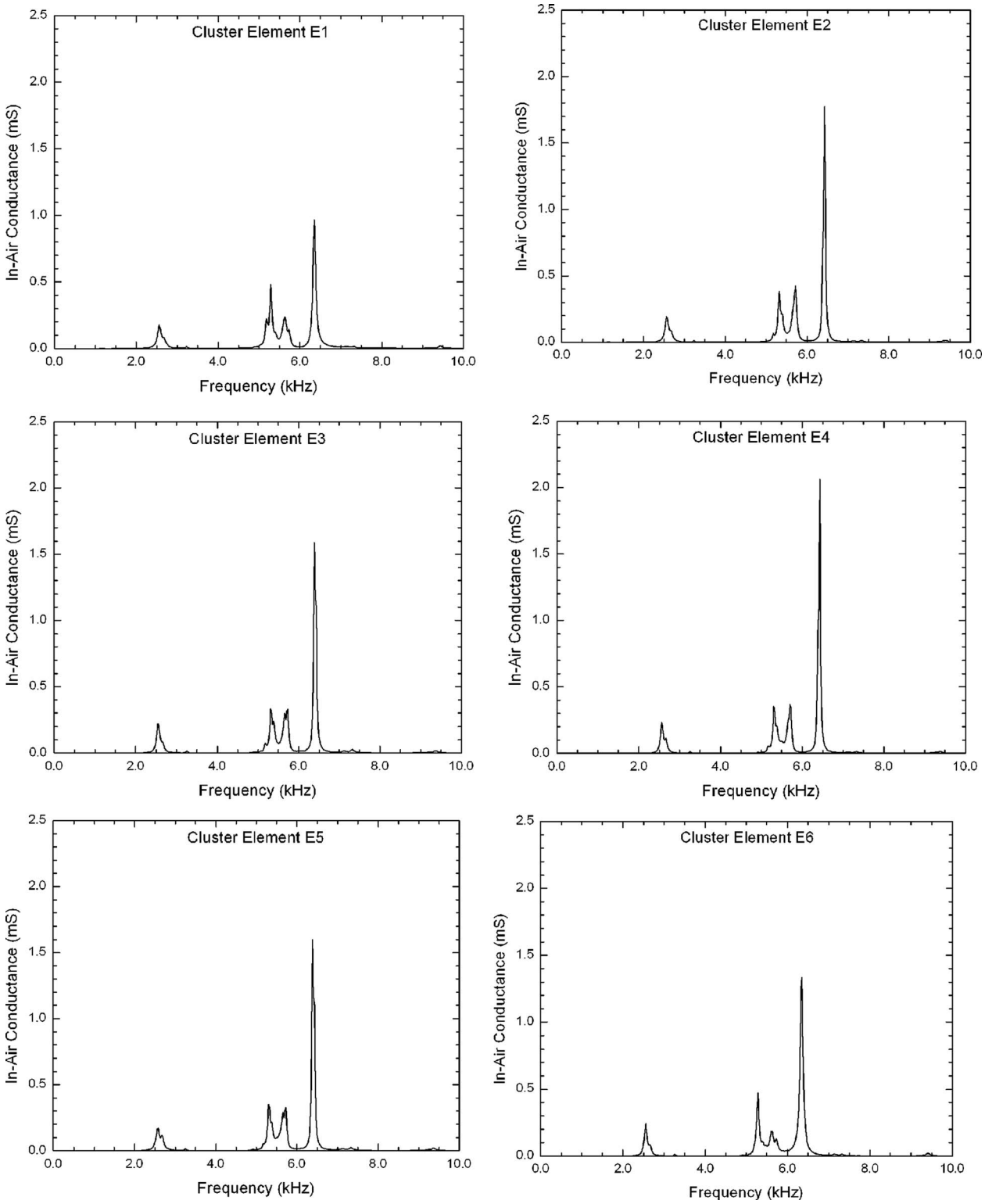
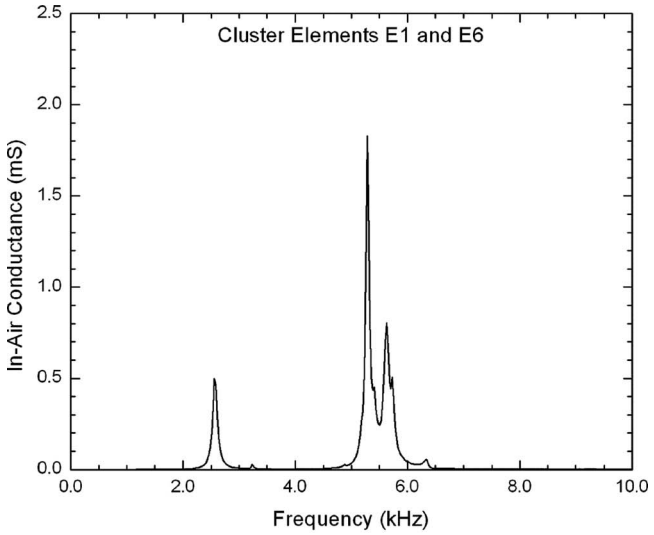


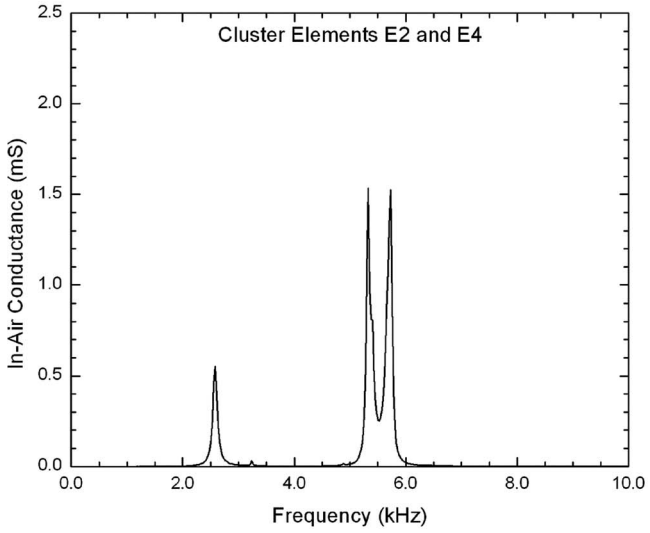
FIG. 7. In-air electrical conductance of the flextensional cluster transducer when driven one element at a time (mass-loaded class I drive).

B. Class III barrel-stave flextensional transducer

The class III barrel-stave flextensional transducer shown in Fig. 3 and at the bottom of Fig. 1 is symmetric at both ends, but twice as long as the class I transducer with similar staves. Therefore, the fundamental flexural resonance frequencies f_0 of the two transducers should be similar. From Fig. 6 and Table I, f_0 is 2.76 kHz, which is about 5% higher than that of the class I transducer. Note that G_0 is higher for the class III transducer (4.37 mS versus 1.16 mS) because there is twice the volume of piezoceramic in the class III design. Owing to its dual-shell design, the class III barrel-stave transducer has a second flexural mode f_1 at 5.78 kHz, as shown in Fig. 6 and listed in Table I. The conductance value G_1 at this resonance peak is 3.08 mS.

Since the class III barrel-stave transducer is longer than the class I transducer, its symmetric longitudinal mode reso-





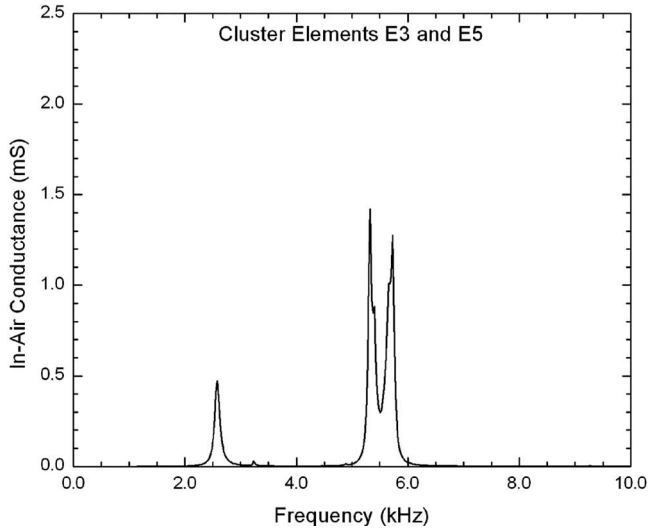


FIG. 8. In-air electrical conductance of the flextensional cluster transducer when driven in collinear pairs (class III drive). Note that E1-E6 is the asymmetric collinear pair.

nance frequency f_S is lower. This is evident in Fig. 6, where the symmetric longitudinal resonance frequency of the class III transducer is 6.25 kHz, which is 29% lower than 8.77 kHz for the class I transducer. With more piezoceramic volume than the class I transducer, G_S for the class III transducer is also higher at 20.1 mS versus 13.2 mS.

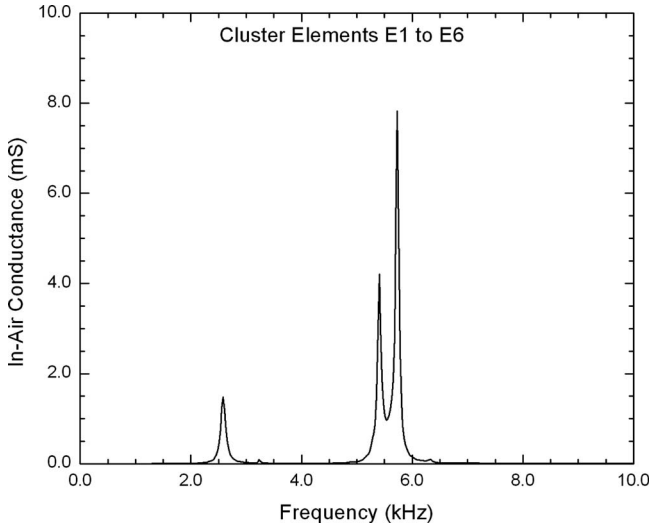


FIG. 9. In-air electrical conductance of the flextensional cluster transducer when all six elements are driven in phase.

C. Broadband cluster transducer

The in-air electrical conductance measurements on the flextensional cluster transducer are shown in Figs. 7–9 and summarized in Table I. These measurements involved the excitation of (1) individual mass-loaded single barrel-stave elements E1–E6; (2) three collinear element pairs E1 and E6, E2 and E4, and E3 and E5; and finally, (3) all six cluster elements simultaneously.

1. Mass-loaded single element excitations

The measured electrical conductance plots for each cluster element excited individually by an applied ac electric field are shown in Fig. 7. In all six cases, a fundamental flexural mode is produced with resonance frequencies f_0 ranging from 2.54 to 2.60 kHz (see Table I). These frequencies are all within 2% of each other and compare well with 2.63 kHz for the class I transducer.

The mass-loaded longitudinal mode in each of the six plots in Fig. 7 has the highest frequency f_M and conductance values G_M. The resonance frequencies are listed in Table I and range from 6.35 to 6.43 kHz, only a 1% variation. Between the low-frequency flexural and high-frequency mass-loaded modes lie two intermediate-frequency longitudinal modes arising from (i) the asymmetric E1-E6 collinear pair (due to the wires exiting the endcap of E1) and (ii) the two symmetric E2-E4 and E3-E5 collinear pairs. As can be seen in the first six rows for the cluster transducer in Table I, the asymmetric resonance frequencies f_A for each of the cluster elements driven alone are lower than the corresponding symmetric resonance frequencies f_S. For example, when element E3 is excited by itself, f_A is 5.30 kHz while f_S is 5.70 kHz. The conductance values associated with these resonance peaks are almost the same at 0.33 and 0.31 mS, respectively. When either element in the asymmetric E1-E6 pair is driven by itself (the top left and bottom right plots in Fig. 7), G_A is

TABLE I. Measured in-air resonance frequencies and conductance values for the flextensional cluster and barrel-stave transducers.

	Flexural modes				Longitudinal modes					
	Fundamental		Second		Asymmetric		Symmetric		Mass loaded	
	f_0 (kHz)	G_0 (mS)	f_1 (kHz)	G_1 (mS)	f_A (kHz)	G_A (mS)	f_S (kHz)	G_S (mS)	f_M (kHz)	G_M (mS)
Transducer (unbooted)										
Class I	2.63	1.16	⋯	⋯	⋯	⋯	8.77	13.2	⋯	⋯
Class III	2.76	4.37	5.78	3.08	⋯	⋯	6.25	20.1	⋯	⋯
Driven cluster elements (booted)										
E1 only	2.60	0.18	⋯	⋯	5.28	0.48	5.63	0.24	6.35	0.97
E2 only	2.56	0.19	⋯	⋯	5.33	0.39	5.73	0.42	6.42	1.77
E3 only	2.54	0.22	⋯	⋯	5.30	0.33	5.70	0.31	6.38	1.59
E4 only	2.56	0.24	⋯	⋯	5.30	0.35	5.70	0.37	6.43	2.06
E5 only	2.58	0.17	⋯	⋯	5.30	0.35	5.73	0.33	6.38	1.60
E6 only	2.56	0.24	⋯	⋯	5.28	0.47	5.60	0.18	6.35	1.34
E1 and E6	2.56	0.50	⋯	⋯	5.28	1.83	5.63	0.81	⋯	⋯
E2 and E4	2.58	0.55	⋯	⋯	5.33	1.53	5.73	1.53	⋯	⋯
E3 and E5	2.58	0.47	⋯	⋯	5.33	1.42	5.73	1.28	⋯	⋯
E1–E6	2.58	1.48	⋯	⋯	5.40	4.21	5.73	7.83	⋯	⋯

significantly higher than G_S. For example, when cluster element E1 is excited, G_A is 0.48 mS whereas G_S is 0.24 mS. Likewise, when element E6 is excited, G_A is 0.47 mS and G_S is 0.18 mS.

2. Collinear pair excitations

Figure 8 shows the measured electrical conductance for the three cases where collinear pairs E1-E6, E2-E4, and E3-E5 were driven by an ac voltage. Note that mass-loaded longitudinal modes did not occur when collinear pairs were driven. Hence, there are only three peaks in each of the plots in Fig. 8, as opposed to four plots in Fig. 7.

From Table I, the resonance frequencies f_0 for the fundamental flexural modes in Fig. 8 were 2.56 kHz for the E1-E6 pair and 2.58 kHz for both the E2-E4 and E3-E5 pairs. These frequencies are essentially the same as those for the six cases where each element was driven individually (see Table I and Fig. 7).

The frequencies f_A and f_S for the asymmetric and symmetric longitudinal modes, whose conductance peaks are prominent in the Fig. 8 plots, are listed in Table I. For the E1-E6 pair, f_A and f_S were 5.28 and 5.63 kHz, respectively, while the same frequencies for the E2-E4 and E3-E5 pairs were 5.33 and 5.73 kHz, respectively. Again, these resonance frequencies compare well to those when only one element in the pair was driven.

Finally, note that the conductance G_A of the asymmetric resonance frequency peak in the top plot in Fig. 8 (E1-E6 pair driven) is more than double the conductance G_S in the same plot, and is consistent with the asymmetric and symmetric peaks when E1 or E6 was driven alone (see top left and bottom right plots in Fig. 7, and the G_A and G_S values listed in Table I). Thus, when either E1 or E6, or both E1 and E6 are driven (these elements are on the asymmetric axis of the cluster transducer), the electrical conductance at the asymmetric longitudinal resonance frequency is greater than the conductance at the symmetric longitudinal resonance frequency. When any of the other four elements, E2–E5, are driven alone or in E2-E4 or E3-E5 collinear pairs (oriented along the two symmetric axes of the cluster transducer), the conductance values at f_A and f_S are almost the same (see the four plots corresponding to E2–E5 excitations in Fig. 7, and the two lower plots corresponding to the driven E2-E4 and E3-E5 pairs in Fig. 8).

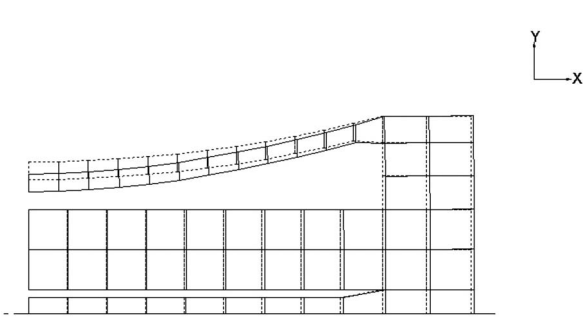


FIG. 10. ATILA™ finite element code-generated (Ref. 21) 2D in-water fundamental flexural mode. Solid lines are displaced transducer.

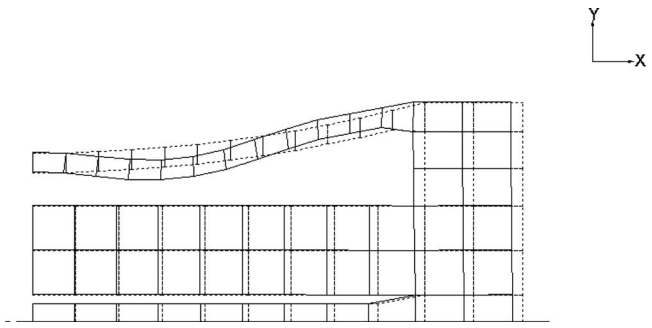


FIG. 11. ATILA™ finite element code-generated (Ref. 21) 2D in-water first longitudinal mode. Solid lines are displaced transducer.

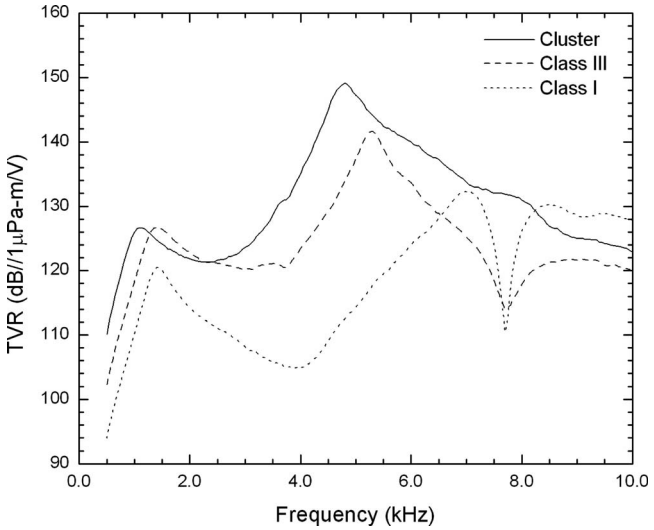


FIG. 12. TVR curves in the radial direction for the class I barrel-stave flextensional transducer (dotted), class III barrel-stave flextensional transducer (dashed), and flextensional cluster transducer with all six elements driven in phase (solid).

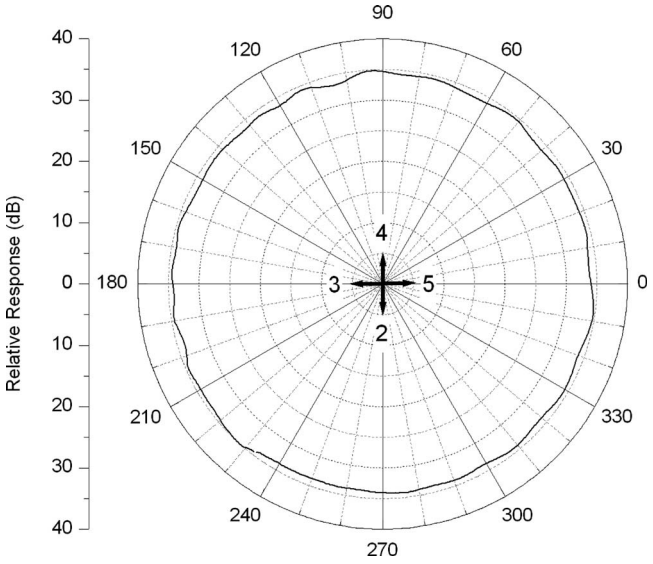


FIG. 13. Directivity pattern for the flextensional cluster transducer at 1.15 kHz (hydrophone at 0°).

3. Simultaneous six-element excitation

When all six cluster elements E1–E6 were driven simultaneously in phase, three resonance peaks were observed below 10 kHz in the measured in-air electrical conductance shown in Fig. 9. The fundamental flexural (f_0), asymmetric longitudinal (f_A), and symmetric longitudinal (f_S) frequencies listed in Table I are 2.58, 5.40, and 5.73 kHz, respectively. These frequencies are similar to those for the single element and collinear pair excitations. Note that the mass-loaded longitudinal mode is not excited when either a collinear pair or all six elements were driven.

V. IN-WATER MEASUREMENTS

The flextensional cluster transducer was tested underwater at a depth of 10 m at the DRDC Atlantic Acoustic Calibration Facility on Bedford Basin, near Halifax, Nova Scotia. All six cluster elements were driven in phase simultaneously with element E1 and the electrical leads pointing upward toward the water surface. Element E6 pointed downward toward the bottom of the basin and the other four elements, E2–E5, were oriented in the horizontal plane with the bottom of E5 pointed toward the reference hydrophone.

The measured TVR of the cluster transducer is shown in Fig. 12, along with the TVR curves for class I and class III barrel-stave flextensional transducers for comparison. Although class I and class III measurements were made at 15 and 30 m water depths, respectively, the effects of hydrostatic pressure from 10 to 30 m on the response data are of the order of 1 dB.

There are three features of interest in the TVR curve in the radial direction for the class III barrel-stave transducer. The low-frequency peak is the fundamental flexural mode, whose resonance frequency and TVR value are 1.45 kHz and 127 dB//1 μPa-m/V. The second resonance peak is the longitudinal mode, which occurs at 5.25 kHz with a TVR of 141 dB//1 μPa-m/V. At 7.75 kHz, there is a dip in the TVR curve caused by out-of-phase coupling between the longitudinal mode at 5.25 kHz, and higher order flexural modes with resonance frequencies greater than 8 kHz.

Because the class I transducer has half the volume of piezoceramic as the class III transducer, its TVR values are lower, as can be seen in Fig. 12. For example, although the fundamental flexural resonance frequencies are the same for both transducers, namely, 1.45 kHz, the class I TVR value at this resonance is 7 dB lower at 120 dB//1 μPa-m/V. Furthermore, since the class I transducer is about half the length of the class III, the longitudinal mode has a higher resonance frequency. Since the geometry of the curved aluminum staves primarily determines the frequencies of the flexural modes, a sharp dip occurs at 7.70 kHz, where the longitudinal and higher order flexural modes couple together. The presence of this dip makes it difficult to estimate the longitudinal resonance frequency, however, on similar class I transducers with staves having the same 20 cm radius of curvature but only 4.83 mm thickness (cf. 5.36 mm for the transducers in this paper), the frequencies of the higher order flexural and longitudinal modes are well separated, with the resonance frequency of the latter at 7.75 kHz, basically where the dip occurs in the class I TVR in Fig. 12.

The TVR for the cluster transducer with all six elements excited simultaneously is also shown in Fig. 12. The fundamental flexural resonance frequency and associated TVR value are 1.15 kHz and 127 dB//1 μPa-m/V, while the longitudinal resonance frequency and TVR value are 4.85 kHz and 149 dB//1 μPa-m/V. Note that there is no dip in the TVR curve above 6 kHz since the longitudinal and higher order flexural modes are well separated. Overall, the cluster TVR resembles more closely that of the class III transducer in the 1–5 kHz band; hence, when all elements are driven in phase, the cluster is behaving acoustically like three orthogonal class III transducers rather than six mass-loaded class I transducers.

The directivity patterns for the cluster transducer were measured at 1.15 and 4.75 kHz with the hydrophone located at 0°, and are shown in Figs. 13 and 14. The orientation of

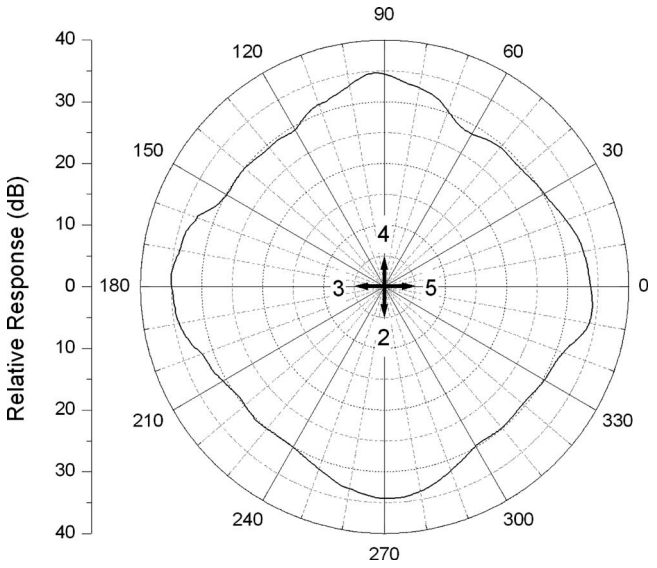


FIG. 14. Directivity pattern for the flextensional cluster transducer at 4.75 kHz (hydrophone at 0°).

the transducer was such that elements E5 and E4 were aligned with the 0° and 90° axes on the polar plots as shown. At 1.15 kHz fundamental flexural resonance frequency, the cluster transducer is omnidirectional when all six elements are driven in phase. At 4.75 kHz near the longitudinal resonance frequency, the directivity pattern in the horizontal plane is almost square with the corners aligned with the ends of the four horizontal elements E2–E5.

VI. CONCLUSIONS

A new flextensional cluster transducer was designed, built, and tested at DRDC Atlantic for naval sonar applications requiring both broadband and directional response capabilities. Electrical conductance measurements performed in air revealed clear low-frequency fundamental flexural modes arising from the six class I barrel-stave elements in the orthogonal cluster transducer structure, whether driven individually, in collinear pairs, or simultaneously. Two longitudinal modes were present for all of these drive conditions as well. The mode with the lowest frequency was identified as the asymmetric longitudinal mode and arose due to the geometrical asymmetry associated with the electrical lead housing on the end of element E1. The higher frequency of the two modes was the usual symmetric longitudinal mode common to all barrel-stave flextensional transducer classes having near geometrical symmetry at both ends. When each of the six elements was driven individually, a mass-loaded longitudinal resonance occurred at frequencies higher than that of the symmetric longitudinal mode. When a collinear pair or all six elements were driven, the mass-loaded mode did not exist.

In water, the six cluster transducer elements were driven in phase and the TVR was measured. The response curve had two resonance peaks in the frequency band of interest—the fundamental flexural resonance at 1.15 kHz and the symmetric longitudinal resonance at 4.85 kHz. Over this band, the cluster transducer's response is qualitatively similar to that of a broadband class III barrel-stave flextensional transducer insofar as the intraresonance response dip is still usable. However, the dip in the response at 7.75 kHz seen in both class I and class III projectors is absent in the cluster's response, giving it more bandwidth with a radial-direction TVR of greater than 120 dB//1 μPa-m/V from 800 Hz to more than 10 kHz (in the alignment measured).

Directivity patterns were measured with all six cluster elements driven in phase at the fundamental flexural resonance frequency 1.15 kHz, and at 4.75 kHz near the longitudinal resonance frequency. The transducer was omnidirectional at the flexural resonance and had a square pattern in the horizontal plane near the longitudinal resonance.

Characterization of this transducer's directivity patterns as a function of frequency and active element combinations is to be carried out with both in-water measurements and finite element analysis. This transducer shows promise in numerous naval applications including broadband communications, multistatic antisubmarine warfare, and broadband low-frequency mine detection. Follow-up work to characterize the broadband cluster transducer's beamsteering capabilities will be conducted and reported later on.

[1]C. J. Purcell and R. A. G. Fleming, "Multi-mode pipe projector," U.S. Patent No. 6,584,039 (24 June 2003).
[2]R. Fleming and J. Fawcett, "Broadband pulse generation using bi-amped MMPP's," Defence Research and Development Canada—Atlantic Technical Memorandum No. 2007-293 (2007).
[3]D. F. Jones and G. W. McMahon, "The design and performance analysis of barrel stave projectors," J. Acoust. Soc. Am. **82**, S75–S75 (1987).
[4]D. F. Jones, "Low-frequency flextensional projectors," in *Proceedings of the Annual Meeting of the Canadian Acoustical Association*, edited by A. J. Cohen (Canadian Acoustical Association, Halifax, Nova Scotia, Canada, 1989), pp. 18–23.
[5]D. F. Jones, "Flextensional barrel-stave projectors," in *Proceedings of the Third International Workshop on Transducers for Sonics and Ultrasonics*, edited by M. D. McCollum, B. F. Hamonic, and O. B. Wilson (Technomic, Lancaster, PA, 1993), pp. 150–159.
[6]D. F. Jones, "Twenty years of barrel-stave flextensional transducer technology in Canada," J. Acoust. Soc. Am. **117**, 2447 (2005).
[7]G. W. McMahon and D. F. Jones, "Barrel stave projector," U.S. Patent No. 4,922,470 (1 May 1990); Canadian Patent No. 1,285,646 (2 July 1991); U.S. Patent Interference No. 102,668 (10 February 1998).
[8]Y. R. Bonin and J. S. Hutton, "Increasing the depth capability of barrel stave projectors," Can. Acoust. **24**, 50 (1996).
[9]D. F. Jones and C. G. Reithmeier, "The acoustic performance of a class III barrel-stave flextensional projector," in *Proceedings of the Undersea Defence Technology Conference* (Nexus Media, Swanley, UK, 1996), pp. 103–108.
[10]G. A. Brigham and L. H. Royster, "Present status in the design of flextensional underwater acoustic transducers," J. Acoust. Soc. Am. **46**, 92 (1969).
[11]L. H. Royster, "The flextensional concept: A new approach to the design of underwater acoustic transducers," Appl. Acoust. **3**, 117–126 (1970).
[12]D. F. Jones, D. J. Lewis, C. G. Reithmeier, and G. A. Brownell, "Barrel-stave flextensional transducers for sonar applications," in *Proceedings of the 1995 ASME Design Engineering Technical Conferences* (American Society of Mechanical Engineers, New York, 1995), Vol. **3**, Pt. B, pp. 517–524.
[13]D. F. Jones and J. F. Lindberg, "Recent transduction developments in Canada and the United States," in *Proceedings of the Institute of Acoustics: Sonar Transducers '95* (Institute of Acoustics, St. Albans, UK, 1995), Vol. **17**, pp. 15–33.
[14]H. B. Miller, "Origin of the 33-driven ceramic ring-stack transducer," J. Acoust. Soc. Am. **86**, 1602–1603 (1989).
[15]H. B. Miller, "Origin of mechanical bias for transducers," J. Acoust. Soc. Am. **35**, 1455 (1963).

[16]H. B. Miller, "Composite electromechanical transducer," U.S. Patent No. 2,930,912 (29 March 1960).
[17]M. B. Moffett and W. L. Clay, Jr., "Demonstration of the power-handling capability of Terfenol-D," J. Acoust. Soc. Am. **93**, 1653–1654 (1993).
[18]H. C. Merchant, "Underwater transducer apparatus," U.S. Patent No. 3,258,738 (28 June 1966).
[19]D. F. Jones, "On the flexural and longitudinal modes of the class I barrel-stave flextensional transducer," Defence Research Establishment Atlantic Technical Memorandum No. 1999-098 (1999).
[20]M. B. Moffett, J. Lindberg, E. A. McLaughlin, and J. M. Powers, "An equivalent circuit model for barrel-stave flextensional transducers," in *Proceedings of the Third Annual Workshop: Transducers for Sonics and Ultrasonics* (Technomic, Lancaster, PA, 1993) pp. 170–180.
[21]B. Hamonic, J. C. Debus, J-N. Decarpigny, D. Boucher, and B. Toucquet, "Analysis of a thin-shell sonar transducer using the finite element method," J. Acoust. Soc. Am. **86**, 1245–1253 (1989).

Model optimization of orthotropic distributed-mode loudspeaker using attached masses

Guochao Lu and Yong Shen
Institute of Acoustics, Key Laboratory of Modern Acoustics (Nanjing University), Ministry of Education, Jiangsu 210093, People's Republic of China

(Received 19 December 2008; revised 31 July 2009; accepted 3 August 2009)

The orthotropic model of the plate is established and the genetic simulated annealing algorithm is developed for optimization of the mode distribution of the orthotropic plate. The experiment results indicate that the orthotropic model can simulate the real plate better. And optimization aimed at the equal distribution of the modes in the orthotropic model is made to improve the corresponding sound pressure responses. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3212943]

PACS number(s): 43.38.Ja, 43.38.Ar [AJZ] Pages: 2294–2300

I. INTRODUCTION

The distributed-mode loudspeaker (DML) has attracted some research interest in recent years,[1–5] for advantages over the cone loudspeaker that include simple structure, wide directivity at high frequencies, and insensitivity to room conditions.

To improve the performance of the DML, various kinds of methods have been proposed: adding a woofer to supply the bass response,[6] optimizing the exciters' positions to improve the directional response,[7] using a porous layer to attain a desired acoustic output,[8] and optimizing with the attached mass method (AMM) to control the vibroacoustic response of a plate.[9–11]

Rebuilding the mode distribution of the plate equally with the AMM to improve the frequency responses is an efficient method.[12] Although the experiment results investigate that the equal mode distribution can improve the frequency response, when in the same conditions, the simulation results of the isotropic model cannot predict the experiment results well.[12] This motivates the development of the simulation model from the isotropic model to the orthotropic one in this paper, in order to make the simulation results correlate the experiments results better.

Partial differential equations (PDEs) of the orthotropic model are based on the Kirchhoff theory of the thin plate and considering the flexural rigidities of the plate are different in the x-axis and y-axis. When neglecting the difference of the flexural rigidities in the x-axis and y-axis, the orthotropic model can be derived to the isotropic one.

Here COMSOL is used for the finite element analysis of the plate, and the surface displacements and the eigenfrequencies of the plate can be derived in COMSOL. The Rayleigh integral is used to calculate the radiated sound pressure level from the baffled plate. Based on the orthotropic model, the genetic simulated annealing algorithm (GSAA) is developed to optimize the mode distribution best. The contrast results of the isotropic model and the orthotropic one verify that the orthotropic model correlates the real plate better, and the optimization results of the orthotropic model show that the optimal design indeed produced better performance in terms of the frequency responses. The results will be discussed and summarized in Sec. V.

II. THEORY

A. PDEs of orthotropic plate vibration

In this paper, the motion and the boundary values of the plate can be specified by the PDEs in the coefficient form for the finite element analysis of the plate, and the geometry is two dimensional.

In detail, the PDE of motion for the orthotropic plate is as follows:

$$D_x\frac{\partial^4 u}{\partial x^4}+2\sqrt{D_xD_y}\frac{\partial^2}{\partial x^2}\frac{\partial^2}{\partial y^2}u+D_y\frac{\partial^4 u}{\partial y^4}+M\frac{\partial^2 u}{\partial t^2}=p, \tag{1}$$

where u is the normal displacement, M is the mass per unit surface area of the plate, and p is the applied pressure acting on the plate in the positive normal direction. Furthermore, D_x is the flexural rigidity of the orthotropic plate in the x-axis, and D_y in the y-axis,

$$D_x=\frac{E_xh^3}{12(1-\nu_x\nu_y)},\quad D_y=\frac{E_yh^3}{12(1-\nu_x\nu_y)}, \tag{2}$$

and E is Young's modulus, ν is Poisson's ratio, h is the thickness of the plate, and the subscripts denote in the x-axis or in the y-axis.

Equation (1) does not include any damping, which does occur in practice. The internal damping of the plate material can be included by taking a complex stiffness of the plate,

$$D_x=d_{0x}(1+j\eta_p),\quad D_y=d_{0y}(1+j\eta_p), \tag{3}$$

where η_p represents the internal damping of the plate.

When $D_x=D_y=D$, the orthotropic model becomes the isotropic one, and the PDE of the isotropic plate can deduce from Eq. (1):

$$D\nabla^4 u+M\frac{\partial^2 u}{\partial t^2}=p, \tag{4}$$

where $\nabla^4=((\partial^2/\partial x^2)+(\partial^2/\partial y^2))^2$ is the biharmonic operator.

The boundary values of the plate are considered as simply supported conditions, so the PDEs of the boundary conditions are as follows:

$$u=0,$$

$$\frac{\partial^2}{\partial x^2}u = 0,$$

$$\frac{\partial^2}{\partial y^2}u = 0. \tag{5}$$

In this paper, the attached masses are considered as the area masses whose thickness can be neglected, so the attached masses do not affect the characteristics of the plate but the area density. After attached with the area masses, the surface of the plate can be separated into two kinds of subdomains. The first subdomain is without the attached masses, whose area density M in Eq. (1) is the plate area density, and the second subdomain is with the attached mass, whose area density M in Eq. (1) is the total area density of the plate and the area mass.

1. Model using a system of one stationary PDE in coefficient form

The coefficient form stationary PDEs to work out the surface displacements of the steady state object are as follows:

$$\nabla \cdot (-c\nabla u - \alpha u + \gamma) + au + \beta \cdot \nabla u = f \quad \text{in} \quad \Omega, \tag{6a}$$

$$n \cdot (c\nabla u + \alpha u - \gamma) + qu = g - h^T \mu, \quad hu = r \quad \text{in} \quad \partial\Omega, \tag{6b}$$

$$n \cdot (c\nabla u + \alpha u - \gamma) + qu = g \quad \text{in} \quad \partial\Omega, \tag{6c}$$

where Ω is the computational domain. Here it means the surface of the object, $\partial\Omega$, is the domain boundary, the boundary of the object, n, is the outward unit normal vector on $\partial\Omega$, u is the normal displacement, $c\nabla u$ is the diffusive flux, αu is the conservative convective flux, γ is the conservative flux source, $\beta \cdot \nabla u$ is the convection term, au is the absorption term, f is the source term, qu is the boundary absorption term, g is the boundary source term, and h is the coefficient matrix whose default form is the identity matrix.

Equation (6a) is the PDE of motion for the object, and it must be satisfied in Ω. Equations (6b) and (6c) are the boundary conditions and one of them must be satisfied on $\partial\Omega$. Equation (6b) is a Dirichlet boundary condition, and Eq. (6c) is a Neumann boundary condition.

In comparison with Eq. (1), it can be seen that the bending wave equation is a fourth-order partial equation, which cannot be represented directly by Eq. (6a). It is possible to translate the fourth-order partial equation into the second-order form by introducing three new variables:

$$u_1 = u, \quad u_2 = \frac{\partial^2}{\partial x^2}u_1, \quad u_3 = \frac{\partial^2}{\partial y^2}u_1. \tag{7}$$

Considering

$$p = Pe^{j\omega t}, \quad u_i = U_i e^{j\omega t} \quad (i = 1,2,3), \tag{8}$$

where ω is the angular frequency.

Then, in the stationary analysis, the corresponding equations can deduce from Eqs. (1), (7), and (8):

$$D_x\frac{\partial^2}{\partial x^2}U_2 + 2\sqrt{D_x D_y}\frac{\partial^2}{\partial y^2}U_2 + D_y\frac{\partial^2}{\partial y^2}U_3 - M\omega^2 U_1 = P,$$

$$U_2 - \frac{\partial^2}{\partial x^2}U_1 = 0,$$

$$U_3 - \frac{\partial^2}{\partial y^2}U_1 = 0. \tag{9}$$

Simply supported boundary conditions are

$$U_1 = 0,$$

$$U_2 = 0,$$

$$U_3 = 0. \tag{10}$$

Then the corresponding equations (9) and (10) can be written in form of Eqs. (6a) and (6b) as follows:

$$\nabla\begin{bmatrix} D_x U_{2x} & 2HU_{2y} + D_y U_{3y} \\ U_{1x} & 0 \\ 0 & U_{1y} \end{bmatrix} + \begin{bmatrix} -\omega^2 M & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}\begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} = \begin{bmatrix} P \\ 0 \\ 0 \end{bmatrix} \quad \text{in} \quad \Omega, \tag{11}$$

$$\begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} = 0 \quad \text{in} \quad \partial\Omega. \tag{12}$$

Let us specify the coefficients c, α, γ, a, β, f, q, g, h, and r,

$$u = \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix}, \quad \gamma = \begin{bmatrix} D_x U_{2x} & 2HU_{2y} + D_y U_{3y} \\ U_{1x} & 0 \\ 0 & U_{1y} \end{bmatrix},$$

$$a = \begin{bmatrix} -\omega^2 M & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad f = \begin{bmatrix} P \\ 0 \\ 0 \end{bmatrix}, \quad h = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{13}$$

and the other parameters are 0.

Based on the developed model, the normal displacement u of the steady state vibration plate can be derived from [v1,v2,...,vn,pe]=postinterp(fem,e1,e2,...,en,xx,...) in COMSOL, which returns the values v1,v2,...,vn of the expressions e1,e2,...,en in the points xx. Then the corresponding normal velocity v in the evaluation points can be calculated as follows:

$$v = \frac{\partial u}{\partial t}. \tag{14}$$

2. Model using a system of one eigenvalue PDE in coefficient form

To solve the eigenvalue problem, not considering the applied pressure p acting on the plate, the equation can be deduced from Eqs. (1) and (7):

$$M\frac{\partial^2}{\partial t^2}u_1 + \nabla(D_x u_{2x} + 2Hu_{2y} + D_y u_{3y}) = 0,$$

$$u_2 = \frac{\partial^2}{\partial x^2}u_1,$$

$$u_3 = \frac{\partial^2}{\partial y^2}u_1. \tag{15}$$

The time derivative terms are second order, and the corresponding PDE is

$$e_a\frac{\partial^2 u}{\partial t^2} + d_a\frac{\partial u}{\partial t} + \nabla\cdot(-c\nabla u - \alpha u + \gamma) + au + \beta\cdot\nabla u = f, \tag{16}$$

where u is the normal displacement, $e_a(\partial^2 u/\partial t^2)$ is the mass term, $d_a(\partial u/\partial t)$ is the damping term, and the other coefficients are the same as Eq. (6a).

The PDEs of the boundary conditions are the same as Eqs. (6b) and (6c). Then the corresponding equation (15) can be written in the form of Eq. (16),

$$\begin{bmatrix} M & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\frac{\partial^2}{\partial t^2}\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} + \nabla\begin{bmatrix} D_x u_{2x} & 2Hu_{2y} + D_y u_{3y} \\ u_{1x} & 0 \\ 0 & u_{1y} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = 0 \quad \text{in} \quad \Omega. \tag{17}$$

Let us specify the coefficients e_a, d_a, c, α, γ, a, β, and f,

$$u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}, \quad e_a = \begin{bmatrix} M & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\gamma = \begin{bmatrix} D_x u_{2x} & 2Hu_{2y} + D_y u_{3y} \\ u_{1x} & 0 \\ 0 & u_{1y} \end{bmatrix}, \quad a = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \tag{18}$$

and the other parameters are 0.

The simply supported boundary conditions are

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = 0 \quad \text{in} \quad \partial\Omega. \tag{19}$$

The eigenvalue derivative terms are second order, and the corresponding PDE is

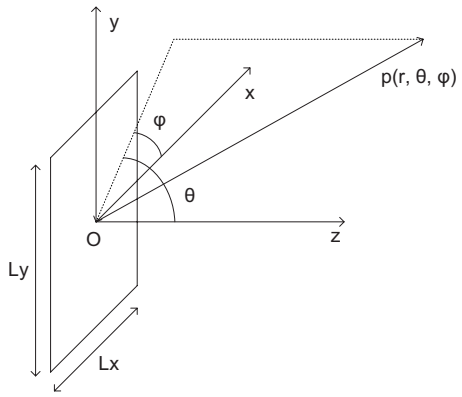


FIG. 1. Coordinate system of the plate for sound radiation simulation analysis.

$$\nabla\cdot(-c\nabla u - \alpha u + \gamma) + au + \beta\cdot\nabla u = d_a\lambda u - e_a\lambda^2 u, \tag{20}$$

where λ is the eigenvalue, and the coefficients are defined the same as Eq. (16). The eigenvalue problem solved in COMSOL is also known as the eigenfrequency problem. The relationship of eigenvalue and eigenfrequency is

$$f = -j\frac{\lambda}{2\pi}. \tag{21}$$

B. Evaluation of sound pressure

In this paper, the radiated sound pressure from the baffled plate can be calculated using the Rayleigh integral,

$$p(r) = j\frac{k\rho_0 c_0}{2\pi}\int\int_s \frac{1}{|r-r_0|}v(r_0)e^{-jk|r-r_0|}dS, \tag{22}$$

where ρ_0 is the air density, c_0 is the speed of sound, k is the wavenumber, r is the field point (x,y,z), r_0 is the source point $(x_0,y_0,0)$, $|r-r_0|$ is the distance between field point and source point, $v(r_0)$ is the normal velocity of the plate at point r_0, and $dS=dx_0dy_0$ is the area of each element at the source point. The integral is taken over the entire area of the plate surface.

Given that the plate is divided into elements, the sound pressure can be calculated with the discrete version of Eq. (22),

$$p(r,\theta,\phi) = j\frac{k\rho_0 c_0}{2\pi}\frac{L_x L_y}{MN}Ev, \tag{23}$$

where $r=\sqrt{(x-x_0)^2+(y-y_0)^2+z^2}$, $(x_0,y_0,0)$ is the central point of the plate, and r,θ,φ are spherical coordinates. The corresponding coordinate system is shown in Fig. 1. L_xL_y is the area of the plate, and MN is the total number of individual elements, which is 100×100 in the simulations of this paper. Furthermore,

$$E = \begin{bmatrix} \exp(-jkr_1)/r_1 \\ \exp(-jkr_2)/r_2 \\ \cdots \\ \exp(-jkr_{MN})/r_{MN} \end{bmatrix}, \tag{24}$$

$$v = [v_1, v_2, \cdots, v_{MN}]^T, \tag{25}$$

where r_n is the distance from the center of the source element $(n=1,2,\ldots,MN)$ to the field point (r,θ,φ), and v is the normal velocity of the plate surface, which can be derived from Eq. (14).

In both the simulations and the experiments, the position of the activating pressure is chosen to be the center of the panel, and the area is 0.02×0.02 m^2. Given the value of the activating force, the activating pressure can be obtained by the equation

$$P=\frac{F}{S}, \tag{26}$$

where F is the activating force and S the area of the activating position. In this study the activating force F is 1 N, and the field point is chosen to be at a distance of 1 m from the panel on the central axis.

C. Optimization model using GSAA

The aim of the optimization is to redistribute the mode frequencies of the plate, reduce the degeneracy of the mode distribution, and consequently improve the sound pressure response of the DML.

Define an index as follows:

$$f=\frac{\left(\frac{1}{N_{\delta f}}\sum \delta f_k\right)^2}{\frac{1}{N_{\delta f}}\sum \delta f_k^2}, \tag{27}$$

where δf_k denote $N_{\delta f}$ individual spacings of eigenfrequencies.

Maximizing the value of f means to reduce the degeneracy. In order to quantify the impact of optimization on the radiated acoustic spectra, define an objective function as follows:

$$\varphi_{\mathrm{SPL}}=\sqrt{\frac{\Sigma(\mathrm{SPL}_{f_k}-\mathrm{SPL}_0)^2}{N_{f_k}}}, \tag{28}$$

where SPL_0 is the mean sound pressure lever in the optimization frequency range, f_k denote the center frequencies of one-24th octave, SPL_{f_k} denote the sound pressure level at f_k, and N_{f_k} denotes the number of f_k.

The smaller value of φ_{SPL} means the smoother frequency response curve. In order to quantity the similarity of two different frequency responses, the objective function can be defined in a similar form:

$$\varphi_{a-b}=\sqrt{\frac{\Sigma(\mathrm{SPL}_{f_k a}-\mathrm{SPL}_{f_k b})^2}{N_{f_k}}}, \tag{29}$$

where f_k denote the center frequencies of 1/24 octave, $\mathrm{SPL}_{f_k i}(i=a,b)$ denote the sound pressure level at f_k, and N_{f_k} denotes the number of f_k.

The smaller value of φ_{a-b} means that two different frequency responses correlate each other better.

In the results, "exp" is short for experiment, which means the results are from the experiments, "iso" is short for isotropic, which means the results are from the simulation of the isotropic model, and "ortho" is short for orthotropic, the same as iso.

The GSAA is used in this optimization, which combines genetic algorithms and the simulated annealing. This hybrid algorithm is capable of overcoming the premature convergence of genetic algorithms and escaping from local optimal solutions.

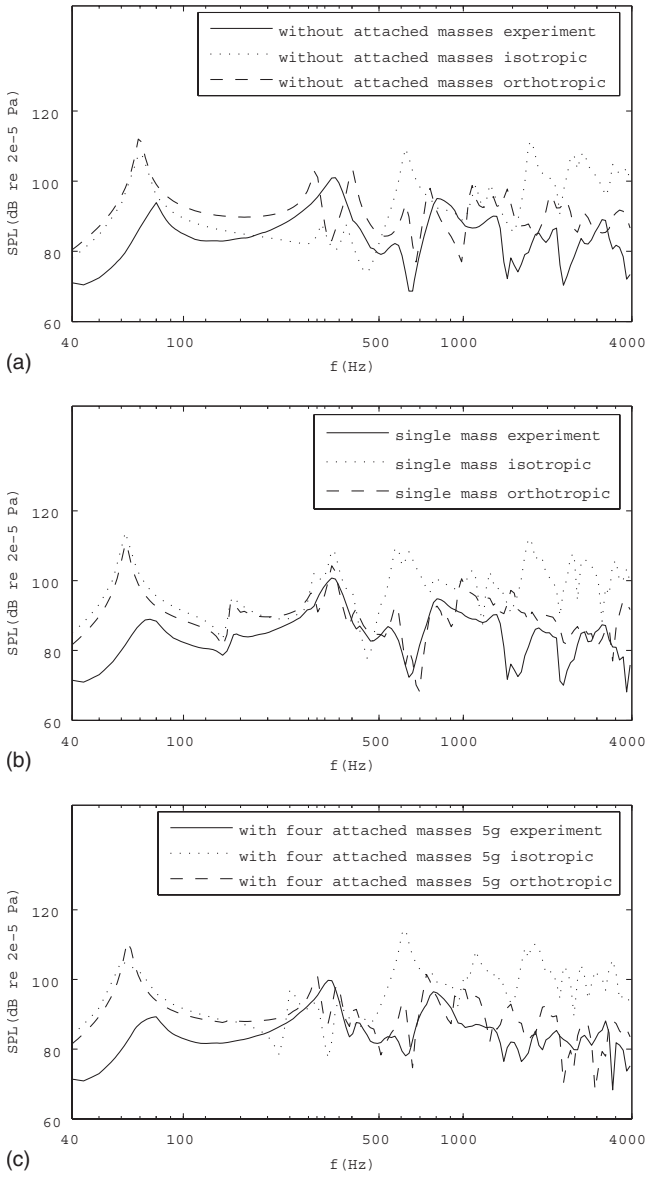


FIG. 2. Contrast of simulations of the isotropic model and orthotropic model with the experiments results. (a) Without attached masses. (b) With one single attached mass; the area of attached mass is 0.02×0.02 m^2, and the weight is 20 g. (c) With four attached masses; the area of attached mass is 0.01×0.01 m^2, and the weight is 5 g.

III. CONTRAST OF ORTHOTROPIC MODEL AND ISOTROPIC MODEL

The simulation results of the orthotropic model and isotropic model are compared with the experiment results which were carried out in an anechoic chamber,[12] as shown in Fig. 2. The parameters of the DML used in the orthotropic model,

TABLE I. Parameters of the DML.

Parameters	
Plate flexural rigidities	$D_x=18.296$ N m
	$D_y=14.282$ N m
Area density	$M=0.697$ kg/m^2
Dimensions	$L_x \times L_y=0.442\times 0.5$ m^2
Damping	$\eta_p=0.05$

isotropic model, and experiments are listed in Table I, and let $D=\sqrt{D_x D_y}$ in the isotropic model.

The comparison results are shown in Tables II and III. Some conclusion can be drawn from these results. First, the AMM works well in improving the sound pressure responses. As shown in Table II, the experiment results and the simulation results of both isotropic model and orthotropic model indicate the improvements of the sound pressure responses after using the AMM. Second, the orthotropic model simulation correlates better with the experiment result than the isotropic model. As shown in Table III, the orthotropic model results can indicate the actual sound pressure responses of the DML better than the isotropic model, especially in the low frequency range, from 40 to 100 Hz, and the high frequency range, from 1000 to 4000 Hz.

IV. OPTIMIZATION SIMULATION RESULTS USING THE ORTHOTROPIC MODEL

The parameters of the DML in optimization using the orthotropic model are the same as listed in Table I, and the value of the objective function is 0.643 while the plate is without attached masses. In the orthotropic model, the numbers, the area, and the positions of the attached masses are considered as the optimized options. The maximum total weight of the attached masses is some 15% of the plate mass, approximately 22.5 g, so the physical effect is that of a plate with slight density variation. The shape of the attached mass is chosen to be square for convenience. The optimization frequency range is 40–4000 Hz.

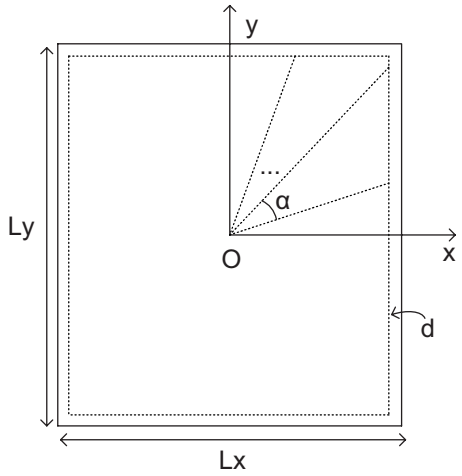


FIG. 3. Area inside of the plate is divide equally into N individual parts according to the number N of the attached masses. The angle of each part α is $2\pi/N$ and the margin between the dashed lines and the axis d is 0.02 m.

When more than one attached mass is optimized, this may result in an overlapping area. To solve the problem, the searching space is divided equally into N individual parts according to the number N of the attached masses, as shown in Fig. 3. The angle of each part α is $2\pi/N$ and the margin between the dashed lines and the axis d is 0.02 m.

In the simulations, more significant digits of the parameters may get a higher value of objective function, but it is not practical for use or experiments. So the parameters and the variables in this paper are calculated with three significant digits.

TABLE II. φ_{SPL} of experiment, isotropic, and orthotropic with various attached masses. The optimization frequency range is 40–4000 Hz.

Number[a]	Weight (g)	Area (m^2)	φ_{SPL}(exp) [b]	φ_{SPL}(iso)	φ_{SPL}(ortho)
None	0	0	7.53	7.82	7.06
Single	20 each	0.02×0.02 each	6.77	7.58	6.83
Four	5 each	0.01×0.01 each	6.54	7.43	6.41

[a]Number means the number of the attached masses, the same as weight and area.
[b]φ_{SPL}(exp) means φ_{SPL} of the experiment frequency responses, the same as φ_{SPL}(iso) and φ_{SPL}(ortho).

TABLE III. Contrast results between the experiments and the isotropic model and between the experiments and the orthotropic model.

Frequency range (Hz)			φ_{a-b}(exp−iso) [a]			φ_{a-b}(exp−ortho)		
Number[b]	Weight (g)	Area (m^2)	40–4000	40–100	1000–4000	40–4000	40–100	1000–4000
None	0	0	8.61	5.64	8.48	5.45	5.42	5.90
Single	20 each	0.02×0.02 each	8.71	7.21	7.82	6.00	7.04	6.13
Four	5 each	0.01×0.01 each	8.18	5.78	5.67	5.45	5.51	4.20

[a]φ_{a-b}(exp−iso) means the contrast between the results of the experiments and the isotropic model, the same as φ_{a-b}(exp−ortho).
[b]Number means the number of the attached masses, the same as weight and area.

TABLE IV. Optimization results of various numbers of the attached masses. The area of attached mass is 0.01 m×0.01 m, and the weight is 5 g.

Number[a]	2	3	4
Positions (x,y)	[0.134 0.065]	[0.004 −0.125]	[0.171 0.0811]
	[−0.156 0.095]	[0.174 0.055]	[–0.127 0.0992]
		[−0.166 0.095]	[−0.151 −0.0405]
			[0.168 −0.102]
Total weight (g)	10.0	15.0	20.0
f	0.809	0.844	0.874

[a]Number means the number of the attached masses, the same as positions and total weight.

The results of optimization are listed in Tables IV and V. Some conclusions can be drawn from the results. First, the value of the objective function is much higher when the plate is optimized with the attached masses. For example, when the plate is attached to four 5 g masses (the area is 10.0 ×10.0 mm^2), the value of the objective function is 0.874, which is quite an improvement compared to 0.643, the value of the plate without attached masses. Second, the result of the optimization is getting better as the number of the attached masses increases. As shown in Table IV, the result of the plate with four attached masses is 0.874, which is much higher than 0.809, the result of the plate with two attached masses. Third, the larger area of the attached mass also improves the value of the objective function, as shown in Table V. When the plate with the larger area attached masses (the area is 10×10 mm^2), the value of the objective function is 0.874, which is better than 0.839, the result of the plate with the smaller area attached masses (the area is 7.1×7.1 mm^2). Fourth, when the plate attached with the same total weight of masses, the more number of the attached masses will improve the value of objective function better. As shown in Tables IV and V, when the total weight of the attached masses is given, the result of the plate with four attached masses (the total weight is 10.0 g, and the area is 7.1 ×7.1 mm^2) is 0.839, which is better than 0.809, the result of the plate with two attached masses (the total weight is 10.0 g, and the area is 10×10 mm^2).

Parts of the corresponding mode frequencies of the optimized plate (the objective value is 0.874) and the plate without attached masses are shown in Fig. 4(a), mode frequencies of the optimized plate are shown in Fig. 4(b), and the corresponding frequency responses are shown in Fig. 5.

TABLE V. Optimization results of various areas of the attached masses. The number of the attached masses is 4 and the area density is 50 kg/m^2. The shape of the attached mass is square, and the area equals the square of the side length.

Side length[a] (mm)	7.1	8.7	10
Positions (x,y)	[0.180 0.152]	[0.0615 0.203]	[0.171 0.0811]
	[0.143 −0.108]	[0.0113 −0.128]	[−0.127 0.0992]
	[−0.146 −0.0218]	[−0.177 −0.151]	[−0.151 −0.0405]
	[−0.177 0.0904]	[−0.143 0.107]	[0.168 −0.102]
Total weight	10.0	15.1	20.0
f	0.839	0.863	0.874

[a]Side length means the side length of the attached masses, the same as positions and total weight.

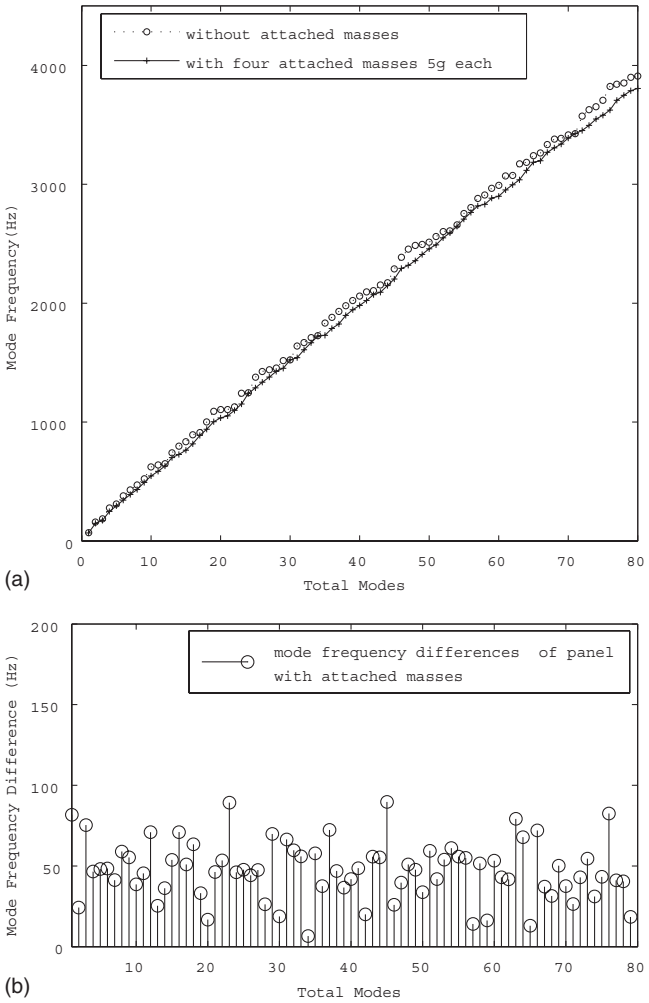


FIG. 4. The area of attached mass is 0.01×0.01 m^2, and the weight is 5 g. (a) Mode frequencies of the plate with and without attached masses. (b) Mode frequency differences of the plate with attached masses.

Comparing the optimal and original simulation results, as shown in Fig. 5 and Table VI, it is obvious that the frequency response is improved. For example, the deep trough between 300 and 400 Hz disappeared, and the curve around 1000 Hz is much smoother after the optimization.

V. CONCLUSIONS

The principal outcome of this work can be summarized in two aspects. First, the orthotropic model of the DML is

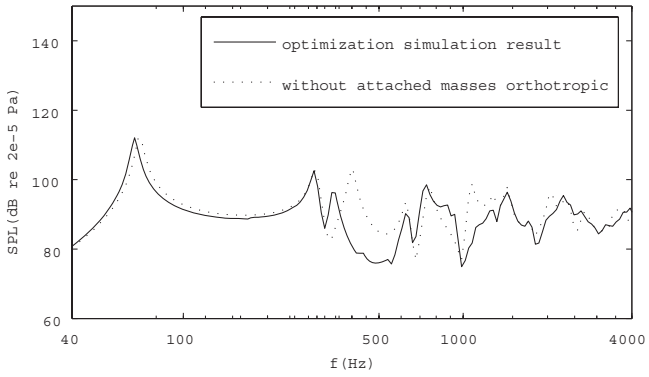


FIG. 5. Simulated sound pressure responses of the plate with and without attached masses. The area of attached mass 0.01×0.01 m^2, and the weight is 5 g.

TABLE VI. Contrast results between the original plane without attached masses and the optimized plane with the attached masses (the objective value is 0.874). The number of the attached masses is 4, the area is 0.01 $\times 0.01$ m^2, and the weight is 5 g and the positions are listed in Table IV.

Frequency range (Hz)	40–4000	300–400	800–1200
φ_{SPL}(original)[a]	7.06	7.00	6.04
φ_{SPL}(optimized)	6.19	4.83	4.82

[a]φ_{SPL}(original) means φ_{SPL} of the original frequency responses, the same as φ_{SPL}(optimized).

implemented in COMSOL. Second, the GSAA procedure is improved to optimize the mode distribution of the DML. Comparing the simulation results and the experiments results, the orthotropic model is more accurate than the isotropic model. In the optimization using the GSAA procedure, a higher value of objective function has been achieved, and the corresponding frequency responses are improved.

In this work, the boundary conditions are considered as simply supported for convenience, and it can be more complex in COMSOL. Although the frequency responses of the DML are improved through the mode distribution optimization, they are still not as smooth as expected. Further research will focus on the relationship between the mode distribution and the frequency response to improve the frequency response of the DML.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (NSFC) under Project No. NSFC 10774075.

[1]M. Roberts, J. Grieco, and C. Ellis, "Diffuse field radiators in automotive sound system design," in Audio Engineering Society 108th Convention, Paris, France (2000).
[2]H. Azima, J. Panzer, and D. Reynaga, "Distributed-mode loudspeakers (DML) in small enclosures," in Audio Engineering Society 106th Convention, Munich, Germany (1999).
[3]V. P. Gontcharov, N. P. Hill, and V. J. Taylor, "Measurement aspects of distributed mode loudspeakers," in Audio Engineering Society 111th Convention, New York (1999).
[4]N. Harris and M. Hawksford, "Spatial bandwidth of diffuse radiation in distributed-mode loudspeakers," in Audio Engineering Society 106th Convention, Munich, Germany (2001).
[5]J. A. Angus, "Distributed mode loudspeaker radiation mechanics," in Audio Engineering Society 108th Convention, Paris, France (2000).
[6]M. R. Bai and T. Huang, "Development of panel loudspeaker system: Design, evaluation and enhancement," J. Acoust. Soc. Am. **109**, 2751–2761 (2001).
[7]M. R. Bai and B. Liu, "Determination of optimal exciter deployment for panel speakers using the genetic algorithm," J. Sound Vib. **269**, 727–743 (2004).
[8]E. Yu. Prokofieva, K. V. Horoshenkov, and N. Harris, "The acoustic emission of a distributed mode loudspeaker near a porous layer," J. Acoust. Soc. Am. **111**, 2665–2670 (2002).
[9]A. J. McMillan and A. J. Keane, "Shifting resonances from a frequency band by applying concentrated masses to a thin rectangular plate," J. Sound Vib. **192**, 549–562 (1996).
[10]A. J. McMillan and A. J. Keane, "Vibration isolation in a thin rectangular plate using a large number of optimally positioned point masses," J. Sound Vib. **202**, 219–234 (1997).
[11]A. Ratle and A. Berry, "Use of genetic algorithms for the vibroacoustic optimization of a plate carrying point-masses," J. Acoust. Soc. Am. **104**, 3385–3397 (1998).
[12]S. Zhang, Y. Shen, X. Shen, and J. Zhou, "Model optimization of distributed mode loudspeaker using attached masses," J. Audio Eng. Soc. **54**, 295–305 (2006).

Modified Škvor/Starr approach in the mechanical-thermal noise analysis of condenser microphone

Chee Wee Tan
Micromachines Centre, School of Mechanical and Aerospace Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore and Center for Environmental Sensing and Modeling (CENSAM) IRG, Singapore-MIT Alliance for Research and Technology (SMART) Centre, 3 Science Drive 2, Singapore 117543, Singapore

Jianmin Miao[a)]
Micromachines Centre, School of Mechanical and Aerospace Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore

(Received 27 February 2009; revised 15 July 2009; accepted 3 August 2009)

Simple analytical expressions of mechanical resistance, such as those formulated by Škvor/Starr, are widely used to describe the mechanical-thermal noise performance of a condenser microphone. However, the Škvor/Starr approach does not consider the location effect of acoustic holes in the backplate and overestimates the total equivalent mechanical resistance and mechanical-thermal noise. In this paper, a modified form of the Škvor/Starr approach is proposed to address this hole location dependent effect. A mode shape factor, which consists of the zero order Bessel and modified Bessel functions, is included in Škvor's mechanical resistance formulation to consider the effect of the hole location in the backplate. With reference to two B&K microphones, the theoretical results of the *A*-weighted mechanical-thermal noise obtained by the modified Škvor/Starr approach are in good agreements with those reported experimental ones.
 [DOI: 10.1121/1.3212917]

PACS number(s): 43.38.Kb, 43.38.Bs, 43.38.Ar [AJZ] Pages: 2301–2305

I. INTRODUCTION

Microelectromechanical system (MEMS) condenser microphones are popular acoustic transducers due to high sensitivity, low power consumption, excellent stability, and flat frequency response.[1,2] The mechanical structure of the condenser microphone comprises of a pair of electrodes that behaves like a variable capacitor in response to external pressure variations. The sensing electrode consists of a deformable elastic diaphragm, while the rigid perforated backplate electrode weakens the air damping effect, which ensures a flat frequency response and reduces the mechanical-thermal noise.[3]

The minute movement of miniaturized mechanical components is adversely affected by molecular thermal agitation, which subsequently gives rise to the mechanical-thermal noise. If this mechanical-thermal noise is not adequately controlled, it can put a limiting factor on the noise performance of many small-sensor systems. The physical origin of the mechanical-thermal noise lies in dissipation mechanisms such as the damping in springs, air damping between two parallel surfaces, and viscous drag in fluids. For the condenser microphone, the dissipation mechanism is represented by the viscous damping losses in the air gap, and slot and holes in the backplate. The mechanical-thermal noise, together with other sources of background noise,[4–6] establishes the lowest limit of acoustic pressure[7,8] that can be picked up by a microphone. In addition, the dissipation mechanism also plays an important role in defining the frequency response characteristics of the microphone. Therefore, an optimized viscous damping value has its origin in a well-designed backplate structure.[9]

A common modeling approach to evaluate the performance of a microphone is the lumped-element method.[10–15] In this approach, through the use of simple analytical expressions, mass, compliance, and damping have their equivalent electrical counterparts in inductance, capacitance, and resistance, respectively. The lumped-element method, as exemplified by the analysis of Gabrielson,[3] has been very popular with many authors.[10–15] In his approach, the total resistance is approximated by a parallel combination of the equivalent mechanical resistances of squeeze-film damping[16] and holes in the backplate.[17] Although the mechanical resistance expression is easy to apply, it does have its inherent limitations. The expression for the mechanical resistance of holes in the backplate only emphasizes the strong dependence on the total open area of holes, number of holes, and air gap thickness, while the positional effect of holes on the mechanical resistance is not considered and properly addressed. An accurate but more complicated modeling approach was described by Zuckerwar,[18,19] and verified in Refs. 18–20 with very good accuracy. In this comprehensive approach, an accurate air resistance term, due to mechanical losses in the air gap, slot, and holes, is also provided.

The approach of Škvor/Starr is relatively straightforward and places a huge emphasis on the air gap thickness (the equivalent mechanical resistance is inversely proportional to it to the power of three). The next important parameter is the

[a)]Author to whom correspondence should be addressed. Electronic mail: mjmmiao@ntu.edu.sg

percentage of open area of holes in the backplate. Thus, the Škvor/Starr approach is based on the concepts of air gap thickness and total hole area and does not take into any considerations of the location effect of holes on the mechanical resistance. However, in the analysis of Zuckerwar, the excited diaphragm deflection profile is taken into consideration by the use of a simple trial function for the diaphragm displacement and the extensive usage of Bessel functions in his formulation. Beside the air gap thickness and total hole area, he also emphasizes the dependency of the air resistance on the location of holes, backplate thickness, and backchamber volume. The omission of the above-mentioned parameters in the analysis of Škvor/Starr probably explains why the approach of Zuckerwar yields a lower, but more accurate, air resistance value than that of Škvor/Starr, a conclusion with which Gabrielson agreed.

In this paper, the effect of the hole location on the mechanical-thermal noise is addressed by making modifications to the mechanical resistance expression of Škvor. A mode shape factor, which is based on the zero order Bessel and modified Bessel functions, is included in Škvor's formulation to consider the effect of the hole location in the backplate. With reference to two B&K microphones, the modified Škvor/Starr approach is compared with other approaches and experimental data.

II. MODIFIED ŠKVOR/STARR APPROACH

Figure 1 illustrates the nomenclature of a condenser microphone with a circular diaphragm of radius a. For two parallel plates, the equivalent mechanical resistance due to squeeze-film damping[16] is given by

$$R_{\text{airgap}} = \frac{3\mu}{2\pi h^3} \quad (\text{N s/m}^5), \tag{1}$$

where μ is the absolute viscosity of air (17.9×10^{-6} N s/m^2) and h is the unpolarized air gap thickness.

The lowest mode shape for a clamped circular diaphragm plate, normalized to a unit displacement at the center of the plate,[21] can be given by

$$W(a_r) = 0.947 J_0(ka_r) + 0.053 I_0(ka_r), \tag{2}$$

where a_r is the radius ring that defines the location of holes in the backplate, $k=3.197/a$ is for the clamped circular diaphragm plate,[21] ka_r is a nondimensional radial coordinate, and $J_0(ka_r)$ and $I_0(ka_r)$ are the zero order Bessel and modified Bessel functions of ka_r, respectively. For regular noncircular diaphragm shapes, such as square, hexagon, and octagon, they can be approximated by a circular one with an equivalent radius, having the same resonant frequency. It must be pointed out that the application of the polarization voltage induces a static deflection of the diaphragm, which, when combined with an incident pressure, alters the deflection mode. As such, for small air gap transducers, this can have a significant impact on the damping characteristics and must be taken into consideration. Figure 2 illustrates the normalized mode shape amplitude for the clamped circular diaphragm plate.

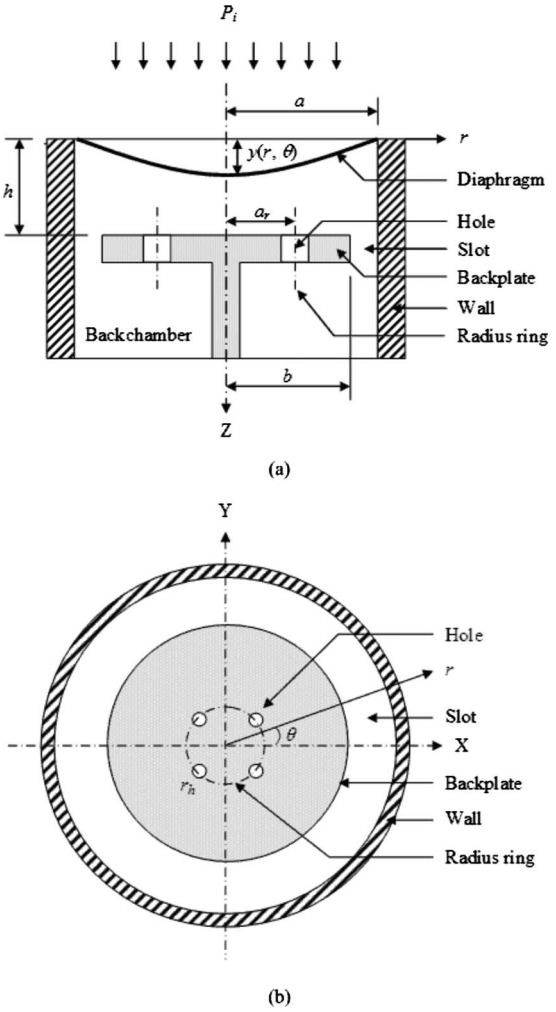


FIG. 1. Nomenclature of a condenser microphone: (a) cross-sectional view and (b) arrangement of acoustic slot and holes in the backplate.

When holes are perforated in the backplate, the viscous damping can be reduced significantly, as air has an alternative path to escape through the holes rather than being squeezed out from the edge of the backplate. In this case, the mechanical resistance corresponding to a single hole[17] is given by

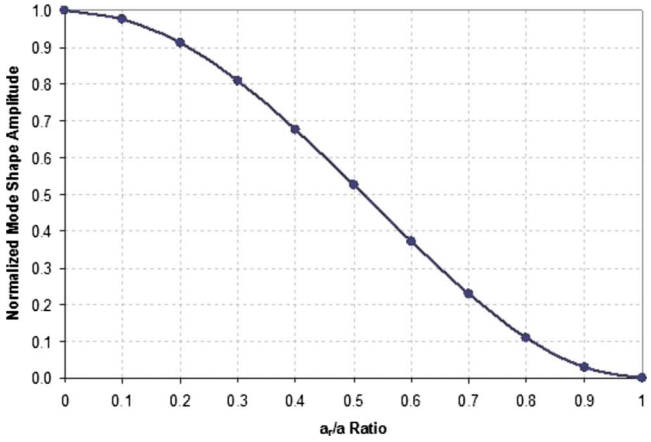


FIG. 2. (Color online) Normalized mode shape amplitude for the clamped circular diaphragm plate versus normalized radius a_r/a.

TABLE I. Parameters of the B&K 0.5 in. (type 4134) microphone.

Parameter	Symbol	Value
Diaphragm radius	a	4.45 mm
Backplate radius	b	3.61 mm
Backplate thickness (at hole location)	l	0.8 mm
Unpolarized air gap	h	20 μm
Acoustic holes		
Number of holes per radius ring	N	6
Location of radius ring	a_r	2 mm
Hole radius	r_h	0.51 mm

$$R_{1\text{hole}} = \frac{12\mu}{\pi h^3} G(A) \quad (\text{N s/m}^5), \tag{3}$$

where $G(A)$ is given by

$$G(A) = \left[\frac{A}{2} - \frac{A^2}{8} - \frac{\ln A}{4} - \frac{3}{8} \right], \tag{4}$$

where A is the ratio of the area of each hole to the area of the diaphragm corresponding to it. The diaphragm area that corresponds to each hole can be approximated by dividing the total diaphragm area by the total number of holes in the backplate. However, for surface-micromachined planar microstructures, in which the air gap and backplate thicknesses are comparable, the hole resistance[22] must be considered and added to the squeeze-film damping. For N perforated holes arranged in a single radius ring, the mechanical resistance can be given by

$$R_{\text{perf}} = \frac{R_{1\text{hole}}}{N} W^2(a_r) \quad (\text{N s/m}^5), \tag{5}$$

where $W(a_r)$ is the mode shape factor that describes the hole location in the backplate.[23] A smaller radius ring, with a corresponding larger mode shape factor, has more dominant contribution to the total equivalent mechanical resistance as compared to a bigger one, with a corresponding smaller mode shape factor, because the diaphragm velocity decreases away from the center of the diaphragm. For multiple radius rings of holes, the total mechanical resistance can be approximated by a parallel combination of the mechanical resistances that are due to each radius ring. Then, the total equivalent air resistance, which accounts for the viscous damping losses in the air gap, slot, and holes, can be approximated by the parallel combination of Eqs. (1) and (5).

Thus, the pressure spectral density of the mechanical-thermal noise can be expressed by

$$P_f = \sqrt{4k_B T (R_{\text{airgap}} \| R_{\text{perf}})}, \tag{6}$$

where k_B is the Boltzmann constant $(1.38 \times 10^{-23}$ J/K) and T is the absolute temperature (K). The background noise pressure level (A-weighted) of the mechanical-thermal noise can be expressed by

$$N_T = \sqrt{\int_{f_1}^{f_2} 4k_B T R A^2(f) df}, \tag{7}$$

where R is the total equivalent air resistance of the air gap, slot, and holes, $A(f)$ is the function of the A-weighted filter, and f_1 and f_2 are 10 Hz and 20 kHz, respectively.

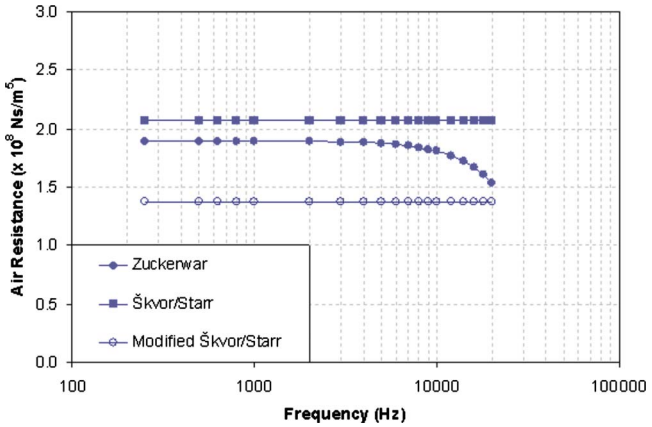


FIG. 3. (Color online) Comparison of approaches for the computation of air resistance versus frequency for the B&K 0.5 in. (type 4134) microphone.

III. RESULTS AND DISCUSSIONS

Table I tabulates the parameters of the B&K 0.5 in. (type 4134) microphone. As listed in Table I, there are a total of six holes arranged in a single radius ring in the backplate. Figure 3 and Table II compare the air resistance and A-weighted mechanical-thermal noise of the B&K 0.5 in. (type 4134) microphone from various sources. As illustrated in Fig. 3, the analysis of Zuckerwar reveals that the air resistance is weakly dependent on the frequency, while both the Škvor/Starr and modified Škvor/Starr approaches are independent of frequency, which are evidenced from their formulations as described in Sec. II. From Fig. 3 and Table II, it is clear that the Škvor/Starr approach overestimates the air resistance and mechanical-thermal noise since it provides the highest air resistance $(2.07 \times 10^8$ N s/m$^5)$ and mechanical-thermal noise [20.2 dB(A)] values among all data sources. However, on the other hand, the modified Škvor/Starr approach provides a mechanical-thermal noise value of 18.3 dB(A), which is much closer to the specification of 18.0 dB(A).

Table III tabulates the parameters of the B&K MEMS microphone.[25] As listed in Table III, four holes are arranged in a single radius ring in the backplate. Figure 4 and Table IV

TABLE II. Comparison of the air resistance and A-weighted mechanical-thermal noise for the B&K 0.5 in. (type 4134) microphone.

Source	$R(\times 10^8$ N s/m$^5)$	Mechanical-thermal noise[a] [dB(A)]
Škvor/Starr[b]	2.07	20.2
Zuckerwar[c]	1.89[d]	19.7
Tarnow[e]	1.54	18.9
Ngo[b]	1.25	18.0
Modified Škvor/Starr	1.35	18.3
Specification[f]	⋯	18.0

[a]Calculated using Eq. (7).
[b]Reference 3.
[c]Reference 18.
[d]Theoretical value at 250 Hz.
[e]Reference 8.
[f]Reference 24.

TABLE III. Parameters of the B&K MEMS microphone.

Parameter	Symbol	Value
Diaphragm radius	a	1.95 mm
Backplate radius	b	1.4 mm
Backplate thickness	l	150 μm
Unpolarized air gap	h	20 μm
Acoustic holes		
Number of holes per radius ring	N	4
Location of radius ring	a_r	0.55 mm
Hole radius	r_h	40 μm

compare the air resistance and A-weighted mechanical-thermal noise of the B&K MEMS microphone. Similarly, as illustrated in Fig. 4, the approach of Zuckerwar reveals the weak dependence of the air resistance on the frequency. From Fig. 4 and Table IV, it is again clear that the Škvor/Starr approach overestimates the air resistance and mechanical-thermal noise, as it provides the highest air resistance (7.25×10^8 N s/m^5) and mechanical-thermal noise [25.6 dB(A)] values among all three approaches. Although the modified Škvor/Starr approach provides a smaller value for the air resistance (6.69×10^8 N s/m^5) and mechanical-thermal noise [25.3 dB(A)], it is still not in good agreement with the measured value of 23.0 dB(A).

As described in Sec. II, the diaphragm area that corresponds to each hole can be approximated by dividing the total diaphragm area by the total number of holes in the backplate. However, in the analysis of Škvor,[17] the center of the hole coincides with that of the collecting diaphragm. As a result, an approximation of the diaphragm area by simply dividing the total diaphragm area by the total number of holes may not be feasible for some cases. For the B&K 0.5 in. (type 4134) microphone, the radius ring is 2.0 mm, which is approximately half of the diaphragm radius. Thus, in this case, the diaphragm area approximation does have its validity, which is reflected by the good agreement of the calculated mechanical-thermal noise to its specification. For the case of highly perforated backplates[10–14] with a regular hole pitch, this diaphragm area approximation is also valid as the holes are evenly distributed throughout the backplate.

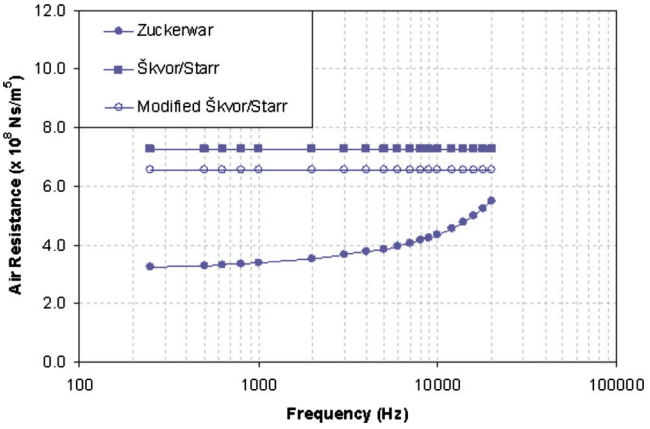


FIG. 4. (Color online) Comparison of approaches for the computation of air resistance versus frequency for the B&K MEMS microphone.

TABLE IV. Comparison of the air resistance and A-weighted mechanical-thermal noise for the B&K MEMS microphone.

Source	$R(\times10^8$ N s/m$^5)$	Mechanical-thermal noise[a] [dB(A)]
Škvor/Starr[b]	7.25	25.6
Zuckerwar[b]	3.22[c]	23.0
Modified Škvor/Starr (area approximation)	6.69	25.3
Modified Škvor/Starr (diameter approximation)	5.65	24.5
Measurement[d]	···	23.0

[a]Calculated using Eq. (7).
[b]Reference 20.
[c]Theoretical value at 250 Hz.
[d]Reference 25.

However, for the B&K MEMS microphone, the radius ring is 0.55 mm, which is less than one-third of the diaphragm radius. Therefore, in this case, the above-mentioned diaphragm area approximation does not have any creditability. As a result, an additional modification has to be made to A and $G(A)$. By equating the circumference of the radius ring to the diameter of four circles, the equivalent radius of the new collecting diaphragm can be estimated. Thus, with a new $G(A)$ value of 0.819, the corrected air resistance is 5.65×10^8 N s/m^5, while the corrected mechanical-thermal noise is 24.5 dB(A), which is in much better agreement with the measured value of 23.0 dB(A) than before. By using the diameter approximation approach, the error between the theoretical and measured mechanical-thermal noise reduces from 10% to 6.5%.

IV. CONCLUSION

The mechanical-thermal noise, together with other sources of background noise, determines the lowest limit of acoustic pressure that can be picked up by a microphone. The simple Škvor/Starr approach is commonly used to determine the mechanical-thermal noise performance of a condenser microphone but it does not consider the location effect of holes in the backplate. To address the hole location effect, a modified form of the Škvor/Starr approach has been proposed, in which a Bessel function-based mode shape factor is incorporated into the mechanical resistance expression of Škvor. With reference to two B&K microphones, the theoretical results obtained by the modified Škvor/Starr approach are in good agreement with those reported experimental ones. This proposed approach can be used to provide a quick and accurate assessment of the mechanical-thermal noise level of the condenser microphone.

[1]P. R. Scheeper, A. G. H. van der Donk, W. Olthuis, and P. Bergveld, "A review of silicon microphones," Sens. Actuators, A **44**, 1–11 (1994).
[2]G. M. Sessler, "Silicon microphones," J. Audio Eng. Soc. **44**, 16–22 (1996).
[3]T. B. Gabrielson, "Mechanical-thermal noise in micromachined acoustic and vibration sensors," IEEE Trans. Electron Devices **40**, 903–909 (1993).
[4]T. B. Gabrielson, "Fundamental noise limits in miniature acoustic and vibration sensors," J. Vibr. Acoust. **117**, 405–410 (1995).
[5]A. J. Zuckerwar, T. R. Kuhn, and R. M. Serbyn, "Background noise in piezoresistive, electret condenser and ceramic microphones," J. Acoust.

Soc. Am. **113**, 3179–3187 (2003).
[6]A. J. Zuckerwar and K. C. T. Ngo, "Measured 1/f noise in the membrane motion of condenser microphones," J. Acoust. Soc. Am. **95**, 1419–1425 (1994).
[7]V. Tarnow, "Thermal noise in microphones and preamplifiers," B&K Technical Review **3**, 3–14 (1972).
[8]V. Tarnow, "The lower limit of detectable sound pressure," J. Acoust. Soc. Am. **82**, 379–381 (1987).
[9]C. W. Tan, Z. H. Wang, J. M. Miao, and X. F. Chen, "A study on the viscous damping effect for diaphragm-based acoustic MEMS applications," J. Micromech. Microeng. **17**, 2253–2263 (2007).
[10]P. R. Scheeper, A. G. H. van der Donk, W. Olthuis, and P. Bergveld, "Fabrication of silicon condenser microphones using single wafer technology," J. Microelectromech. Syst. **1**, 147–154 (1992).
[11]W. Kühnel and G. Hess, "A silicon condenser microphone with structured backplate and silicon nitride membrane," Sens. Actuators, A **30**, 251–258 (1992).
[12]P. R. Scheeper, W. Olthuis, and P. Bergveld, "Improvement of the performance of microphones with a silicon nitride diaphragm and backplate," Sens. Actuators, A **40**, 179–186 (1994).
[13]M. Pedersen, W. Olthuis, and P. Bergveld, "A silicon condenser microphone with polyimide diaphragm and backplate," Sens. Actuators, A **63**, 97–104 (1997).
[14]M. Pedersen, W. Olthuis, and P. Bergveld, "High-performance condenser microphone with fully integrated CMOS amplifier and DC-DC voltage converter," J. Microelectromech. Syst. **7**, 387–394 (1998).
[15]R. Nadal-Guardia, A. M. Brosa, and A. Dehé, "AC transfer function of electrostatic capacitive sensors based on the 1-D equivalent model: Application to silicon microphones," J. Microelectromech. Syst. **12**, 972–978 (2003).
[16]J. B. Starr, "Squeeze-film damping in solid-state accelerometers," in Proceedings of the IEEE Workshop in Solid-State Sensor and Actuator, Fourth Technical Digest (1990), pp. 44–47.
[17]Z. Škvor, "On the acoustical resistance due to viscous losses in the air gap of electrostatic transducers," Acustica **19**, 295–299 (1967).
[18]A. J. Zuckerwar, "Theoretical response of condenser microphones," J. Acoust. Soc. Am. **64**, 1278–1285 (1978).
[19]A. J. Zuckerwar, "Principles of operation of condenser microphones," in *AIP Handbook of Condenser Microphone: Theory, Calibration and Measurements*, edited by G. S. K. Wong and T. F. W. Embleton (AIP, New York, 1995), Chap. 3, pp. 37–69.
[20]C. W. Tan and J. M. Miao, "Analytical modeling for bulk-micromachined condenser microphone," J. Acoust. Soc. Am. **120**, 750–761 (2006).
[21]A. W. Leissa, *Vibration of Plates* (AIP, New York, 1993).
[22]D. Homentcovschi and R. N. Miles, "Modeling of viscous damping of perforated planar microstructures. Applications in acoustics," J. Acoust. Soc. Am. **116**, 2939–2947 (2004).
[23]I. J. Oppenheim, A. Jain, and D. W. Greve, "Electrical characterization of coupled and uncoupled MEMS ultrasonic transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 297–304 (2003).
[24]G. S. K. Wong, "Microphone data applications," in *AIP Handbook of Condenser Microphone: Theory, Calibration and Measurements*, edited by G. S. K. Wong and T. F. W. Embleton (AIP, New York, 1995), Chap. 19, pp. 291–301.
[25]P. R. Scheeper, B. Nordstrand, J. O. Gulløv, B. Liu, T. Clausen, L. Midjord, and T. Storgaard-Larsen, "A new measurement microphone based on MEMS technology," J. Microelectromech. Syst. **12**, 880–891 (2003).

Vibration absorption using non-dissipative complex attachments with impacts and parametric stiffness

N. Roveri
Department of Mechanics and Aeronautics, University of Rome, "La Sapienza," Via Eudossiana, 18, 00184 Rome, Italy

A. Carcaterra[a)]
Department of Mechanics and Aeronautics, University of Rome, "La Sapienza," Via Eudossiana, 18, 00184 Rome, Italy and Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

A. Akay
Department of Mechanical Engineering, Bilkent University, Ankara 06800, Turkey and Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

(Received 31 October 2008; revised 30 July 2009; accepted 31 July 2009)

Studies on prototypical systems that consist of a set of complex attachments, coupled to a primary structure characterized by a single degree of freedom system, have shown that vibratory energy can be transported away from the primary through use of complex undamped resonators. Properties and use of these subsystems as by energy absorbers have also been proposed, particularly using attachments that consist of a large set of resonators. These ideas have been originally developed for linear systems and they provided insight into energy sharing phenomenon in large structures like ships, airplanes, and cars, where interior substructures interact with a master structure, e.g., the hull, the fuselage, or the car body. This paper examines the effects of nonlinearities that develop in the attachments, making them even more complex. Specifically, two different nonlinearities are considered: (1) Those generated by impacts that develop among the attached resonators, and (2) parametric effects produced by time-varying stiffness of the resonators. Both the impacts and the parametric effects improve the results obtained using linear oscillators in terms of inhibiting transported energy from returning to the primary structure. The results are indeed comparable with those obtained using linear oscillators but with special frequency distributions, as in the findings of some recent papers by the same authors. Numerically obtained results show how energy is confined among the attached oscillators. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3212942]

PACS number(s): 43.40.At, 43.40.Kd, 43.40.Jc, 43.40.Tm [ADP] Pages: 2306–2314

I. INTRODUCTION

An extensive literature exists on energy distribution in prototypical systems that consist of a set of linear parallel undamped resonators, called here as the attachment, all connected to a common vibrating structure, and often referred to as the primary or master structure. The pioneering work of Pierce *et al.*[1] investigated a plate with a complex attachment demonstrating its unconventional damping property in the frequency-domain, and in Refs. 2 and 3 the problem is reconsidered, looking at the properties of a prototype master structure with attached set of weakly damped resonators. In Ref. 4, the damping effect produced by this prototypical system is analytically demonstrated, even independently of any local energy dissipation, for an infinite number of resonators and with a particular frequency distribution. The problem was further analyzed, focusing on the temporary nature of the energy storage for a finite number of attached resonators[5] and on the energy redistribution process and equipartition in large undamped resonators.[6] In Ref. 7 the intrinsic properties of attachments are identified, which control the speed of energy sharing between a master and the attachment and the time the energy takes to be transferred back to the master. Several studies examined the conditions that, even in the absence of energy dissipation, prevent energy transport back to the master, which lead to the so called near-irreversibility condition[8–10] also confirmed by experimental tests.[11] Finally, the problem of an efficient design of a multi-degrees of freedom tuned-mass-damper has been also considered in the context of control theory.[12]

In all these studies, energy redistribution process is considered in the framework of (i) linear interaction between the master and the attached resonators and (ii) in the absence of any direct interaction among the resonators, except through their reactions on the primary.

This paper addresses the effects of nonlinearities on energy transport by introducing nonlinear interaction—elastic collisions—among the resonators and a parametric instantaneous variation in the stiffness of the attached oscillators. The motivation for investigating these effects is summarized briefly as follows.

(a) In Ref. 10 it is shown how a damping effect on the

[a)]Author to whom correspondence should be addressed. Electronic mail: a.carcaterra@dma.ing.uniroma1.it

master develops due to the attachment only when the uncoupled natural frequency of the master belongs to the interval B described by the natural frequencies of the oscillators within the attachment. Conversely, if the master frequency falls outside of this bandwidth B, the energy sharing process is inhibited, significantly decoupling the master and the attachment.

(b) Under the conditions of the first point in case (a), most of the energy is transferred form the master to a limited number of resonators, i.e., to those oscillators having their natural frequencies closer to that of the master, thus concentrating the energy over a limited part of the attachment.[7]

(c) Energy is continuously transferred and stored into the attachment for a period of time, but after a characteristic return time,[7] it is transferred back to the master.

These observations naturally lead to investigating means to produce permanent energy storage within the attachment by modification of the linear system.

The behavior described in case (a) suggests that the master and the attachment can be energy-coupled or decoupled by just modifying the characteristic frequency distribution within the attachment during the vibration process as follows. In a linear system initially with a frequency distribution tuned with the master frequency, the energy is transferred from the master to the attachment. Following this transfer, when the condition of energy flow inversion from the attachment to the master becomes imminent (and this condition can be even theoretically predicted as in Ref. 7), the frequencies of the resonators of the attachment are suddenly modified, moving them far away from the master frequency, creating an energy-decoupling condition, and "freezing" the energy within the attachment. This strategy described in Sec. III.

An alternative approach, which amounts to producing an energy spreading effect, is to introduce direct interactions among the resonators within the attachment, permitting to them to have free and direct energy exchange. As described in Sec. II, letting oscillators develop impacts among them redistributes energy from those most energized to the others.

These nonlinear techniques also produce a near-irreversible energy transfer between the master and the attachment similar to that described in Ref. 10 for linear systems but using special frequency distribution within the attachment.

II. IMPACTS WITHIN THE ATTACHMENT

The prototypical two degrees of freedom system that produces impacts between adjacent oscillators, is schematically described in Fig. 1, with m, k_1, k_2, $x_1(t)$, and $x_2(t)$ representing mass, stiffness (k_1,k_2), and displacement of each resonator, respectively. It represents the characteristic module for elastic collision interaction used in the more general attachment investigated ahead, involving indeed multiple resonators, and its preliminary analysis helps in a better understanding of the general case.

The nonlinear behavior emerges as the relative distance $|x_1(t)-x_2(t)|$ equals the gap g and an impact between the

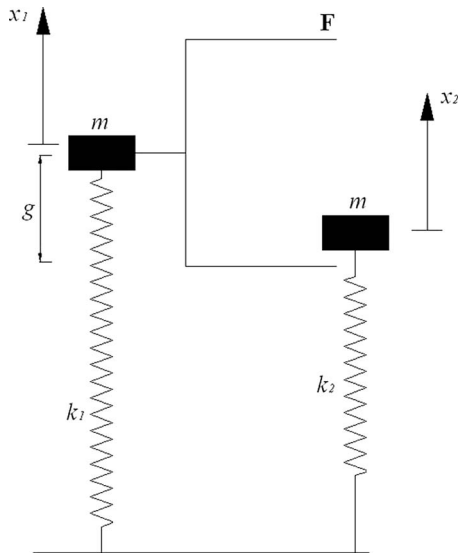


FIG. 1. The two-resonator impact-coupling.

resonators takes place through the impact frame **F**. An assumption of perfect elastic collision is made. The equations of energy and the momentum conservation imply

$$(\dot{x}_1^2(t_+)+\dot{x}_2^2(t_+))\frac{m}{2}=(\dot{x}_1^2(t_-)+\dot{x}_2^2(t_-))\frac{m}{2},$$

$$m(\dot{x}_1(t_+)-\dot{x}_1(t_-))=-m(\dot{x}_2(t_+)-\dot{x}_2(t_-)), \tag{1}$$

where t_- and t_+ are the time just preceding and subsequent to the impact, respectively. It follows

$$\dot{x}_1(t_+)=\dot{x}_2(t_-),$$

$$\dot{x}_2(t_+)=\dot{x}_1(t_-), \tag{2}$$

meaning the resonators just exchange their velocities during an impact. Equation (2) is used to study the impacts within the complete attachment consisting of a plurality of resonators.

Therefore, the complete system represented in Fig. 2, in the absence of external forces, is described by the equations

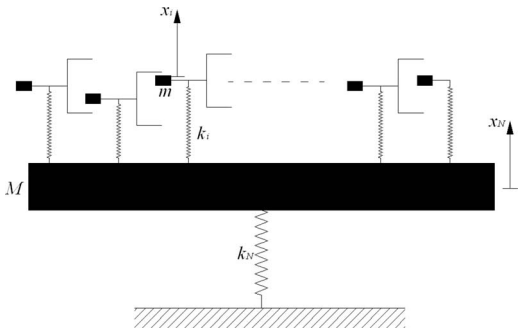


FIG. 2. Master-attachment prototype system.

$$m\ddot{x}_j(t) + k_j(x_j(t) - x_N(t)) = \sum_k I_k^{j,i}\delta(t - t_k), \quad j = 1,2,\ldots,N-1,$$

$$M\ddot{x}_N(t) + k_N x_N(t) + \sum_{j=1}^{N-1} k_j(x_N(t) - x_j(t)) = 0, \tag{3}$$

where the index N designates the master, $1,2,\ldots,N-1$ are used for the oscillators of the attachment, m, k_j, M, k_N, $x_j(t)$, and t are the mass and the stiffness of each oscillator of the attachment, the mass and the stiffness of the master, the displacement of the jth oscillator, and time, respectively, $I_k^{j,i}$ represents the impulse exchanged between the jth and the ith resonators at time t_k, $I_k^{i,j} = -I_k^{j,i}$, and $\delta(t-t_k)$ is the Dirac delta function. However, accordingly with the system depicted in Fig. 2, the elastic collision interactions represented by $I_k^{j,i}$ are restricted to the resonators with the nearest neighbors.

Matrix form for Eq. (3) reads

$$\mathbf{M}\ddot{x} + \mathbf{K}x = f(x,\dot{x}), \tag{4}$$

where $\mathbf{M}$ and $\mathbf{K}$ are the mass and stiffness matrices, and $f(x,\dot{x})$ represents the conservative, internal, and impact forces.

Equation (4) is piecewise linear and an iterative analytic solution at each iteration step can be expressed as

$$x(x_0,\dot{x}_0,t,t_0) = \sum_{r=1}^{N}\left[\mathbf{u}_r{}^T\mathbf{M}x_0\cos(\omega_r t - \omega_r t_0) + \mathbf{u}_r{}^T\mathbf{M}\dot{x}_0\frac{1}{\omega_r}\sin(\omega_r t - \omega_r t_0)\right]\cdot\mathbf{u}_r, \tag{5}$$

where ω_r and $\mathbf{u}_r$ are the eigenfrequency and the corresponding eigenvector, respectively, x_0 and $\dot{x}_0$ represent the initial displacement and velocity at t_0, respectively. Expression (5) is used iteratively to build the solution $s^k(t)$, which is a set of continuous functions for each time interval $[t_k,t_{k+1})$, within which no impact takes place. For $t \in [0,t_1)$ Eq. (5) yields

$$s^0(t) = x(x_0,\dot{x}_0,t,t_0) \quad \forall\, t \in [0,t_1). \tag{6}$$

With the initial conditions at $t_0=0$,

$$x_0 = \begin{Bmatrix} 0 \\ \vdots \\ 0 \end{Bmatrix}; \quad \dot{x}_0 = \begin{Bmatrix} 0 \\ \vdots \\ V_0 \end{Bmatrix}. \tag{7}$$

For $t \in [t_1,t_2)$ Eq. (5) becomes

$$s^1(t) = x(x_0,\dot{x}_0,t,t_0) \quad \forall\, t \in [t_1,t_2) \tag{8}$$

with initial condition on displacement as

$$x_0 = s^0(t_{1^-}). \tag{9}$$

The initial velocities are obtained using Eq. (2) for each impacting pair of resonators j and i at t_1

$$\dot{x}_{0j} = \dot{s}_i^0(t_{1^-})$$

$$\dot{x}_{0i} = \dot{s}_j^0(t_{1^-}) \tag{10}$$

Finally, for each oscillator h that does not undergo an impact,

$$\dot{x}_{0h} = \dot{s}_h^0(t_{1^-}) \tag{11}$$

is the initial condition at $t_0=t_1$.

The computational process starts with Eqs. (6)–(11), iteratively repeated up to the desired end time.

It would be emphasized how this procedure leads to a piecewise continuous solution using linear analysis within time spans between impacts in conjunction with velocity rules, given by Eq. (2), which impose velocity discontinuities on the resonators.

The model represented by Eqs. (1)–(11) is used to describe the energy sharing process between the master and the attachment. In Sec. IV the energy time history of the master $E_N(t) = 1/2M(\dot{x}_N^2 + \omega_M^2 x_N^2)$ and its time average $\lim_{T\to\infty} 1/T\int_0^T E_N(t)dt$, where $\omega_M = \sqrt{k_N/M}$, are considered together with the average energy of the satellite oscillators.

III. PARAMETRIC EFFECTS: TIME-VARYING STIFFNESS

Several previous studies of the linear oscillators have shown how initially imparted energy to a master migrates to the attached oscillators.[2,6,7] In particular, these studies have also shown how special frequency distributions of the oscillators enhance the transport of energy rapidly from the master to the oscillators.[7] Theoretical, numerical, as well as experimental evidences of this phenomenon have been offered in Refs. 7–10. These results show that energy exchange between the master and its satellites takes place through a preferential frequency bandwidth B, as pointed out in case (a) in Sec. I, which must contain the master frequency, while the energy sharing process is inhibited when the master frequency falls outside this bandwidth.

Based on these considerations, the concept proposed here employs parametrically variable stiffness, with instantaneous variations, for the satellite oscillators; after an initial tuning period during which the master frequency falls within B, the satellite frequencies are shifted in a way that the master frequency is left outside B. Thus, the energy sharing process is inhibited before the energy can return to the master, confining the energy permanently within the attachment. Such a system, analogous to the one considered in Sec. II, still behaves linearly in each time interval.

The satellite oscillators all have equal mass m, while their initial stiffness is selected within the set $S \equiv \{k_r, r = 1,\ldots,N-1 \,|\, k_r \neq k_s \text{ for } r \neq s\}$. The initial value of the time-varying stiffness $\chi_i(t)$ of the ith oscillator falls within S.

Of the two approaches proposed here to parametrically vary stiffness, the simpler one uses a time-dependent stiffness $\chi_i(t)$ defined as

$$\chi_i(t) = k_i + \Delta k_i H(t - t^*), \quad \text{and} \quad k_i \in S,$$

$$\sqrt{\frac{k_i + \Delta k_i}{m}} > \sqrt{\frac{k^{\max}}{m}}, \quad \forall\, i, \tag{13}$$

where H is the Heaviside step function and $k^{\max} = \max\{k_i\}$, $\Delta k_i > 0$.

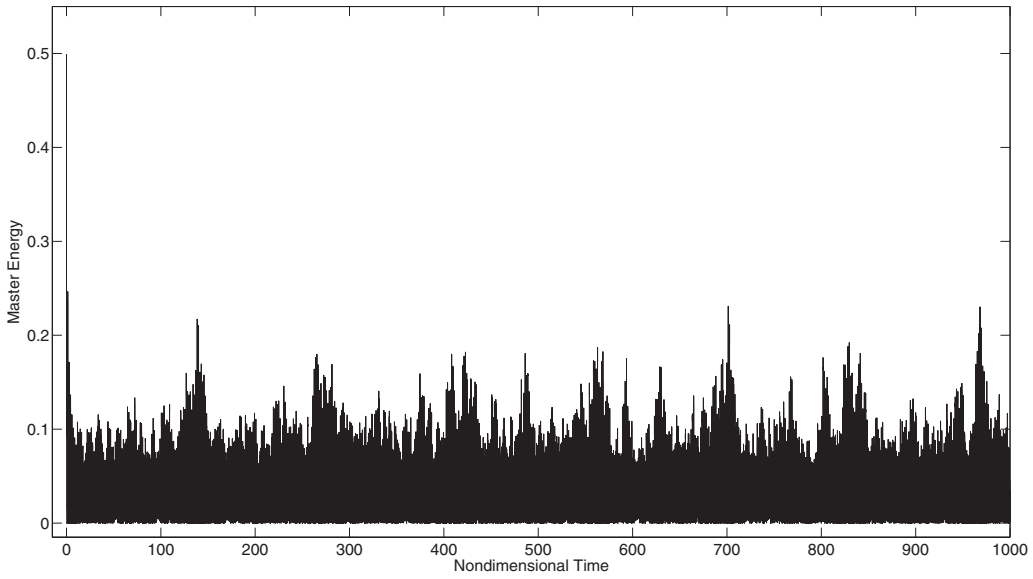


FIG. 3. Energy time history of the master for the linear system.

The initial uncoupled oscillator frequencies $\omega_r = \sqrt{k_r/m}$ all belong to the bandwidth $B \equiv \{0, \sqrt{k^{\max}/m}\}$, which includes the master frequency ω_M.

The stiffness χ_i takes values, after a period t^*, within a set $T \equiv \{k_r + \Delta k_r, r = 1, \ldots, N-1\}$, defined in Eq. (13). Equation (13) implies that for $t \geq t^*$, the all oscillator frequencies moved away from the bandwidth B, thus inhibiting energy sharing between the master and the satellite oscillators beyond time t^*, freezing the energy within the attachment.

Note in this case how, without prescribed values for Δk_i, except as described in Eq. (13), the frequency distribution $\sqrt{k_i + \Delta k_i/m}$ obtained $t > t^*$ differs from the initial one $\sqrt{k_i/m}$, and that the stiffness values χ_i for $t > t^*$ no longer belong to S, i.e., S and T have an empty intersection.

The second strategy for the parametric stiffness control uses the same frequency distribution at all times t, i.e., the stiffness of the attachment always belongs to the same set S at all times. This second procedure follows the steps described below.

(1) The oscillators within the attachment are subdivided into two groups: $R^{(L)}$ and $R^{(H)}$. Those included in $R^{(H)}$ retain most of the total energy (as shown in Sec. IV), the remaining belong to $R^{(L)}$. In general, the number N_L of resonators of $R^{(L)}$ exceeds the number N_H of resonators of $R^{(H)}$.
(2) After time t^*, the stiffnesses $\chi_r^{(H)}(t)$ $(r = 1, \ldots, N_H)$ of the resonators in the group $R^{(H)}$, are simply interchanged with some of the stiffness $\chi_i^{(L)}(t)$ $(i = 1, \ldots, N_L)$ belonging to group $R^{(L)}$. The following expressions express this process formally:

$$\chi_r^{(H)}(t) = k_r + [\chi_s^{(L)}(t) - k_r] H(t - t^*),$$

$$r = 1, 2, \ldots, N_H, \quad s \in \{1, 2, \ldots, N_L\}, \tag{14}$$

$$\chi_s^{(L)}(t) = k_s + [\chi_s^{(H)}(t) - k_s] H(t - t^*).$$

Because of this simple interchange, no new additional frequencies are introduced to the attachment. This implies that $T \equiv S$, meaning the initial and the final frequency distributions within the attachment are the same, even though the stiffness of the individual resonators are changed with time in accordance with Eq. (14).

In the spirit of the present context, the system considered here remains conservative even under stiffness modifications. To achieve this goal, the stiffness variation for the ith oscillator would be introduced when $x(t) - x_i(t) = 0$, such that the perturbation of $\chi_i(t)$ does not modify the potential energy stored in the spring, leaving the total energy of the resonator unchanged. Use of this technique suggests the need to introduce the stiffness modifications at different times for each resonator of the set. In practice, however, it is more convenient to modify the stiffness values simultaneously for all the resonators at the same time t^* without checking their individual position. In order to make the process simpler, the modified spring stiffness for each oscillator k_i^{new} must have the same energy as the original one (stiffness k_i^{old}),

$$\frac{1}{2} k_i^{\text{new}} [x(t^*) - x_i^{\text{new}}(t^*)]^2 = \frac{1}{2} k_i^{\text{old}} [x(t^*) - x_i^{\text{old}}(t^*)]^2,$$

$$x_i^{\text{new}}(t^*) = x(t^*) - \sqrt{\frac{k_i^{\text{old}}}{k_i^{\text{new}}}} [x(t^*) - x_i^{\text{old}}(t^*)].$$

Thus, by modifying the stiffness, the corresponding elongations $x_i^{\text{new}}(t^*)$ of the spring for each oscillator is also modified with respect to its original values $x_i^{\text{old}}(t^*)$, in accordance with the energy conservation requirement expressed above. This condition implies that the energy balance of each oscillator is preserved, but with a different static equilibrium position after the stiffness change.

A final consideration concerns the selection of the time t^*. This is roughly the time it takes for the energy of the

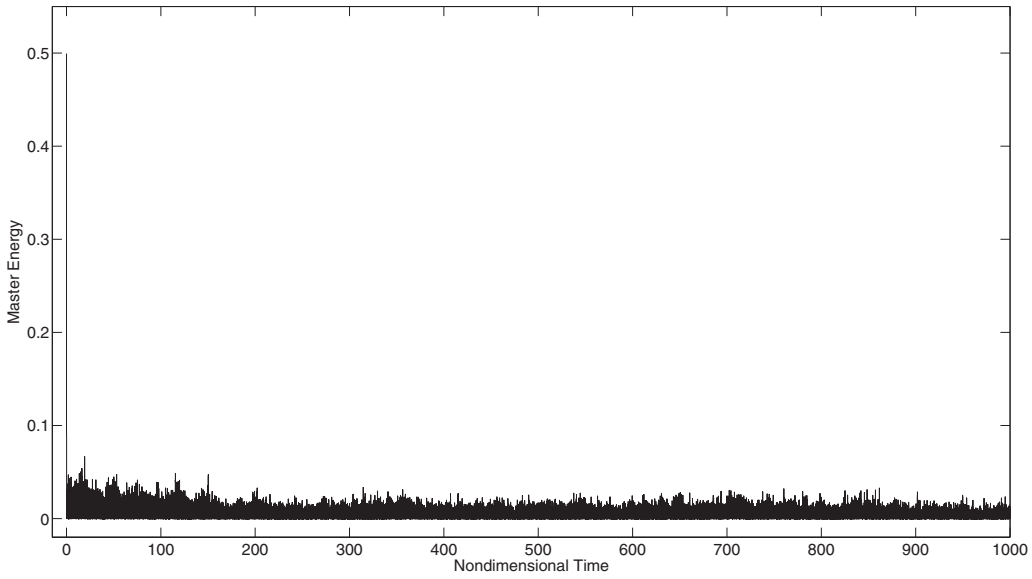


FIG. 4. Energy time history of the master for the nonlinear system.

master to completely migrate to the attachment. As shown in Ref. 7, for a linear attachment, the return time t_{ret} indicates the time after which the energy returns back to the master, and in Ref. 7 it is also shown how it depends on the selected frequency distribution within the attachment and on the total number $N-1$ of resonators. The time periods of t_{ret} and t^* are similar; in fact, as the energy of the master is transferred to the attachment, phase synchronization among the resonators within the attachment takes place and the energy is suddenly returned to the master.

Therefore, time t^* must be long enough to allow the most effective energy transfer from the master to the attachment, but shorter than t_{ret} to avoid the energy reverse process. A suitable choice for t^* could be $t^* \approx 0.9 t_{\text{ret}}$ (the one used in the simulations) so that the return effect is prevented and the energy absorbed from the master is nearly all confined in the resonators of the attachment.

IV. NUMERICAL RESULTS

As a numerical implementation of the model defined in Sec. II, based on elastic collisions, consider a case in which a total of $N=130$ attached oscillators have equally spaced frequencies within the bandwidth $[\omega_M/50, 2\omega_M]$. The mass ratio between the mass of the attachment and the mass of the master is about 0.1, and $m/M=1/(10N)$.

The choice of the characteristic gap g can be operated, following many different criteria. For the present numerical

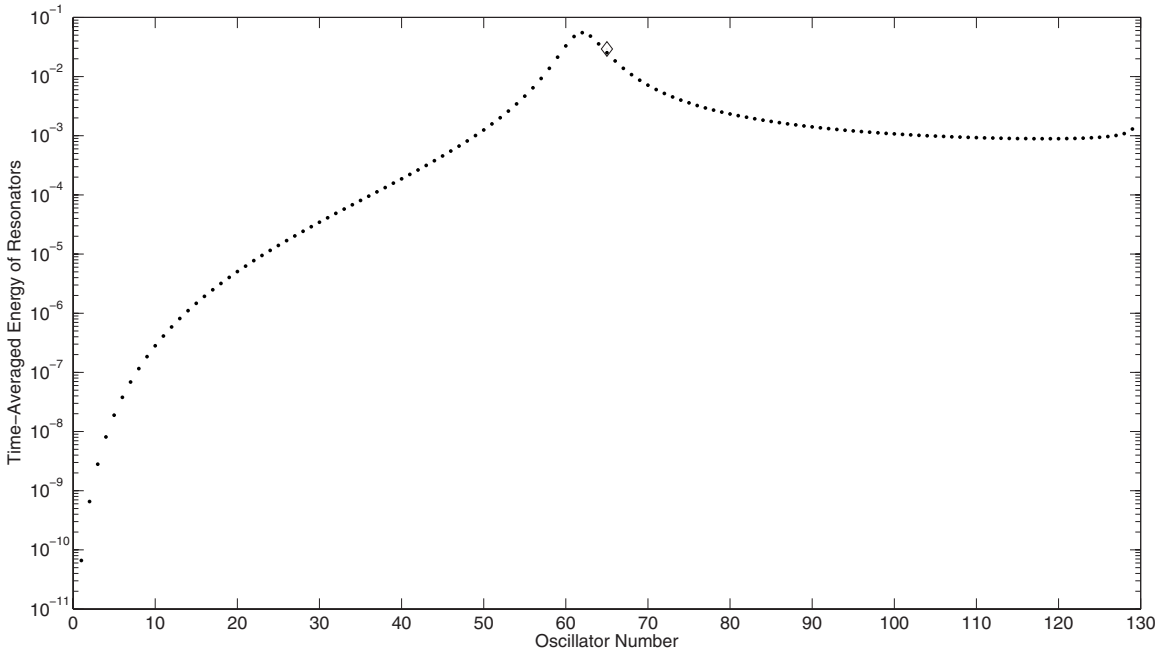


FIG. 5. Time-averaged energy of each resonator for the linear system: dots represent the oscillators energies and the diamond represent the energy of the master; y-axis is log-scale.

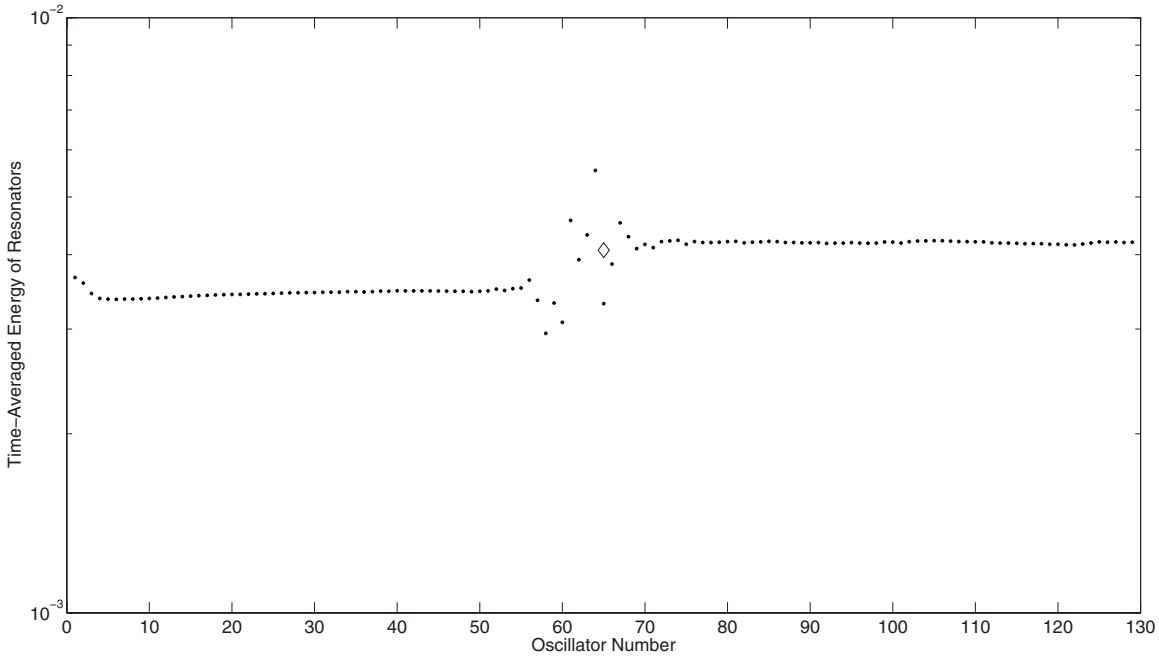


FIG. 6. Time-averaged energy of each resonator for the nonlinear system (symbols as in Fig. 5).

simulations, the steps have been used as follows:

- the system response is simulated, without elastic collisions, during the time interval $[0, 2T_{\max}]$, where $T_{\max}$ is the maximum natural period of the system;
- the distances $d_i(t) = |x_{i+1}(t) - x_i(t)|$, where $i = 1, N-2$, between neighbors resonators are monitored, and their maxima D_i within the time interval $[0, 2T_{\max}]$ are extracted; and
- if $g_{\max} = \max\{D_i, i = 1, N-2\}$ then the gap g is chosen (equal for all the resonator pairs) as a fraction of $g_{\max}$, namely, $g = 0.8 g_{\max}$.

The idea behind this procedure is physically simple: the process of collision is initially activated for those resonators having an energy level close to their maximum. It is empirically found that this criterion produces good results, and a

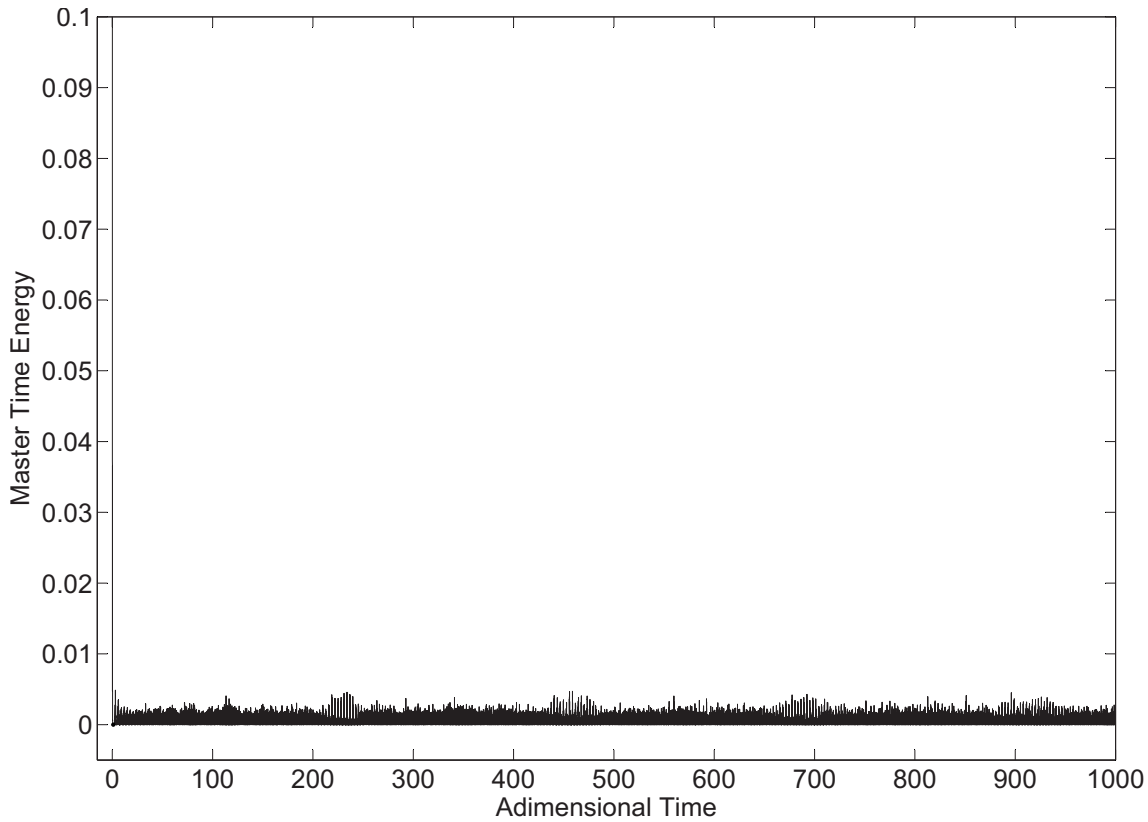


FIG. 7. Energy time history of the master (case a).

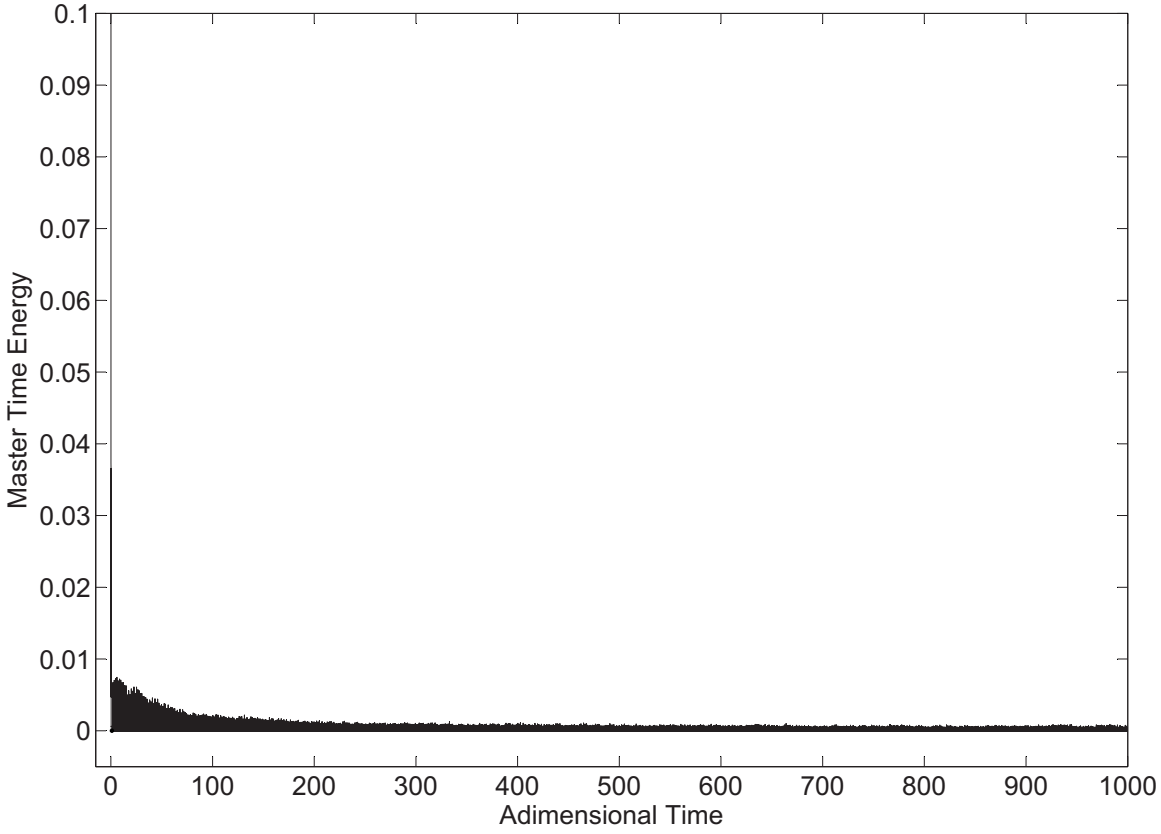


FIG. 8. Energy time history of the master (case b).

systematic analysis of the effect of g on the energy absorption capability of the attachment will be the subject of future investigations.

The total initial energy imparted on the primary is $E_{tot}=0.5$. Energy time histories of the master are plotted in Figs. 3 and 4 for the linear and nonlinear systems, respectively. Time axis is non-dimensional, taking T_{max} as the reference time. Presence of impacts enhances the energy absorption capability of the attachment, significantly reducing the vibration amplitude of the master.

Figures 5 and 6 show the time-averaged energy stored in each resonator for the cases of linear and nonlinear systems, respectively. The time base over which the average is computed is equal to $1\,000^{*}\ T_{max}$. For the linear case the energy

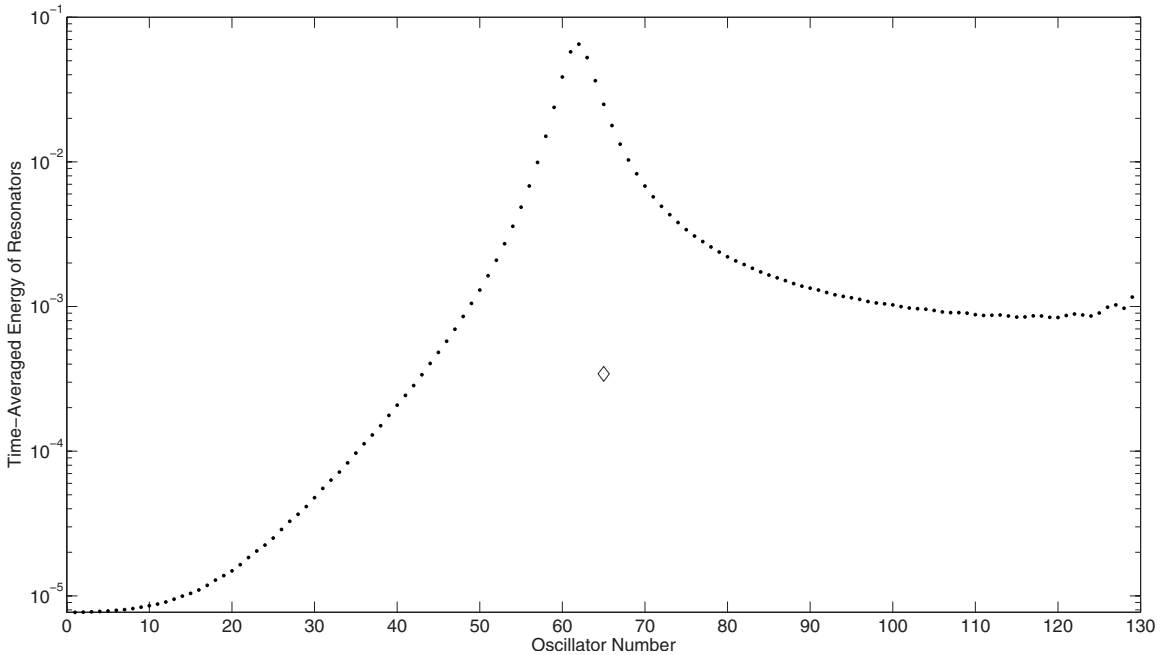


FIG. 9. Time-averaged energy of each resonator (case a) (symbols as in Fig. 5).

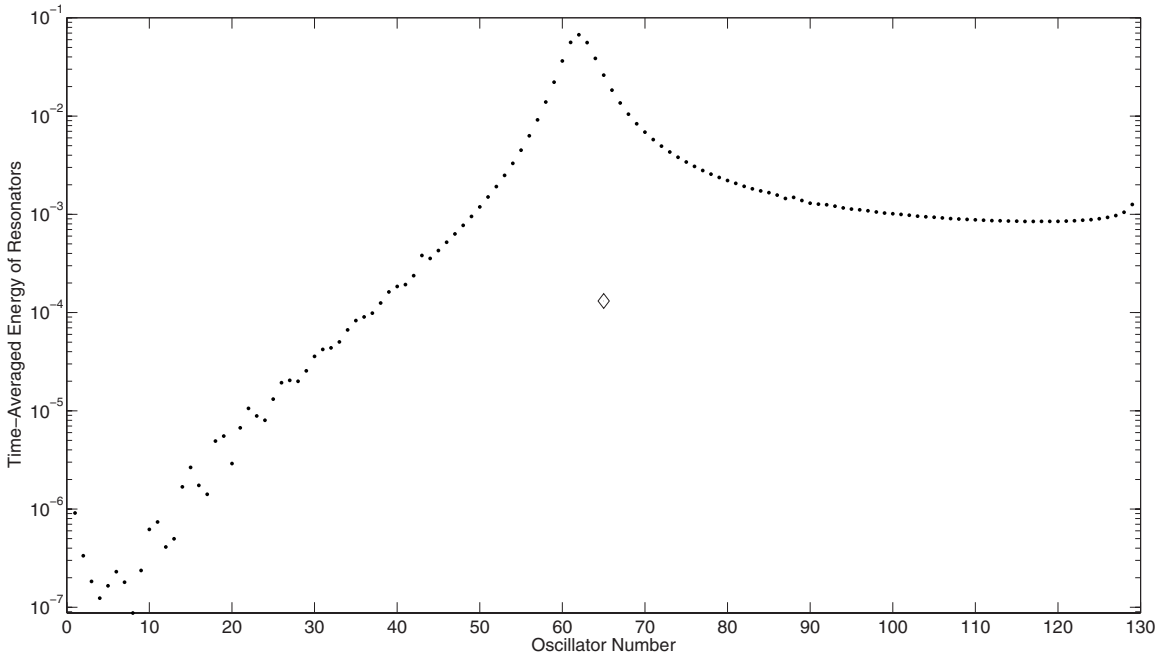


FIG. 10. Time-averaged energy of each resonator (case b) (symbols as in Fig. 5).

is mainly shared among the master and a small group of resonators tuned to the master's frequency, in agreement with the findings of Ref. 7. In fact, in the curve of Fig. 5, a sharp peak appears around the master frequency. In the nonlinear case, energy in the attachment is almost equally shared among resonators, mostly with a value around $E_{\text{tot}}/N \approx 0.0038$, approaching energy equipartitioning.

Turning to the alternative mechanism of introducing nonlinearity through stiffness modification, a system with the same bandwidth and number of degrees of freedom as in the previous case is considered. As shown in Sec. III, the two cases investigated are as follows:

a. parametric control by hardening all the stiffness of the satellite structure, by setting $\Delta k_i = k_{N-1} - k_1$, such that all the frequencies of the satellites for $t > t^*$ fall outside the bandwidth B; and
b. parametric control by using the same frequency distribution.

Energy time histories of the master are shown in Figs. 7 and 8 for cases (a) and (b), respectively.

Comparing these results with those in Fig. 3 demonstrates how effective parametric control can be in making the master time energy approach zero.

Figures 9 and 10 show the time-averaged energy stored into each resonator, again for cases (a) and (b), respectively. In both cases the energy of the master is much lower than the equipartition value $E_{\text{tot}}/N \approx 0.0038$, and the first frequency shift starts when the master energy is very close to zero. Comparing Figs. 9 and 10 with Fig. 6, a very limited spreading of the energy among resonators is observed; the shape of the energy spectra shown in Figs. 9 and 10 is much closer to the one obtained for the linear case, as shown in Fig. 5, than to the one with impacts (Fig. 6).

Figures 4, 7, and 8 show that the master energies are quite similar for both cases, as well as impacts and stiffness modification among resonators of the satellite structure, although they are based on different physical phenomena. As shown in Figs. 6, 9, and 10, in the case of impacts, total energy is equally spread among the resonators, thus the energy of the master is close to E_{tot}/N, while in the case of frequency shifts through stiffness modifications there is no equidistribution; the energy is trapped in a group of satellite resonators, and the master energy remains nearly constant and equal to its value at the time of the first shift.

As a final point, energy equipartition within the attachment can be produced by (i) a nonlinear mechanism, through elastic impacts among the resonators and regardless of the frequency distribution of the system, and (ii) a purely linear mechanism from a proper selection of frequency distribution of the oscillators, as recently shown in Ref. 13.

V. SUMMARY AND CONCLUSIONS

The present paper considers the problem of energy sharing between a master and a plurality of parallel resonators attached to it, introducing two elements of novelty with respect to the previous investigations regarding the presence of collisions among the resonators within the attachment and parametric variation of their stiffness. Purpose of both of these approaches is to make the energy transfer from the primary to the attachment permanent. For both cases, numerical results show very good energy absorption capability of the attachments introduced in this paper. They may be considered as alternatives to selecting special frequency distributions within the attachment[10] to produce a near-irreversible energy transfer from the master to an attached set of linear oscillators. Of note, the results obtained with the techniques described here are not significantly sensitive to

the particular frequency distribution within the attachment. Moreover, the nonlinear mechanisms introduced here produce an energy absorption capability of the attachment that is very similar to that obtained using the previously reported special frequency distribution in linear oscillators. A physical reason that can qualitatively explain this equivalence relies on the ability of the nonlinear mechanisms introduced here in spreading the energy across the resonators of the attachment, an ability that is shared with linear systems having the particular frequency distribution reported in Ref. 10.

[1]A. D. Pierce, V. W. Sparrow, and D. A. Russel, "Fundamental structural-acosutic idealization for structure with fuzzy internals," J. Vibr. Acoust. **117**, 339–348 (1995).

[2]M. Strasberg and D. Feit, "Vibration damping of large structures induced by attached small resonant structures," J. Acoust. Soc. Am. **99**, 335–344 (1996).

[3]G. Maidanik, "Induced damping by a nearly continuous distribution of a nearly undamped oscillators: Linear analysis," J. Sound Vib. **240**, 717–731 (2001).

[4]R. J. Nagem, I. Veljkovic, and G. Sandri, "Vibration damping by a continuous distribution of undamped oscillators," J. Sound Vib. **207**, 429–434 (1997).

[5]R. L. Weaver, "The effect of an undamped finite degree of freedom 'fuzzy' substructure: Numerical solution and theoretical discussion," J. Acoust. Soc. Am. **100**, 3159–3164 (1996).

[6]R. L. Weaver, "Equipartition and mean square response in large undamped structures," J. Acoust. Soc. Am. **110**, 894–903 (2001).

[7]A. Carcaterra and A. Akay, "Transient energy exchange between a primary structure and a set of oscillators: Return time and apparent damping," J. Acoust. Soc. Am. **115**, 683–696 (2004).

[8]I. M. Koç, A. Carcaterra, Z. Xu, and A. Akay, "Energy sinks: Vibration absorption by an optimal set of undamped oscillators," J. Acoust. Soc. Am. **118**, 3031–3042 (2005).

[9]A. Carcaterra, A. Akay, and I. M. Koç, "Near-irreversibility in a conservative linear structure with singularity points in its modal density," J. Acoust. Soc. Am. **119**, 2141–2149 (2006).

[10]A. Carcaterra and A. Akay, "Theoretical foundation of apparent damping and energy irreversible energy exchange in linear conservative dynamical systems," J. Acoust. Soc. Am. **121**, 1971–1982 (2007).

[11]A. Akay, Z. Xu, A. Carcaterra, and I. M. Koç "Experiments on vibration absorption using energy sinks," J. Acoust. Soc. Am. **118**, 3043–3049 (2005).

[12]L. Zuo and S. A. Nayfeh, "Minimax optimization of multi-degrees of freedom tuned-mass-dampers," J. Sound Vib. **272**, 893–908 (2004).

[13]N. Roveri, A. Carcaterra, and A. Akay, "Energy equipartition and frequency distribution in complex attachments," J. Acoust. Soc. Am. **126**, 122–128 (2009).

Bistatic scattering from submerged unexploded ordnance lying on a sediment

J. A. Bucaro,[a] H. Simpson, L. Kraus, L. R. Dragonette, T. Yoder, and B. H. Houston
Naval Research Laboratory, Washington, DC 20375

(Received 5 March 2009; revised 15 July 2009; accepted 30 July 2009)

The broadband bistatic target strengths (TSs) of two submerged unexploded ordnance (UXO) targets have been measured in the NRL sediment pool facility. The targets—a 5 in. rocket and a 155 mm projectile—were among the targets whose monostatic TSs were measured and reported previously by the authors. Bistatic TS measurements were made for 0° (target front) and 90° (target side) incident source directions, and include both backscattered and forward scattered echo angles over a complete 360° with the targets placed proud of the sediment surface. For the two source angles used, each target exhibits two strong highlights: a backscattered specular-like echo and a forward scattered response. The TS levels of the former are shown to agree reasonably well with predictions, based on scattering from rigid disks and cylinders, while the levels of the latter with predictions from radar cross section models, based on simple geometric optics appropriately modified. The bistatic TS levels observed for the proud case provide comparable or higher levels of broadband TS relative to free-field monostatic measurements. It is concluded that access to bistatic echo information in operations aimed at detecting submerged UXO targets could provide an important capability.
[DOI: 10.1121/1.3212920]

PACS number(s): 43.40.Fz, 43.20.Fn, 43.20.Gp, 43.30.Xm [DF] Pages: 2315–2323

I. INTRODUCTION

Many active and former military installations have ordnance ranges and training areas with adjacent water environments in which unexploded ordnance (UXO) now exists, due to wartime activities, dumping, and accidents. Over time, such geographic areas are becoming less and less remote as the adjacent lands become further developed, and the potential hazard to the public from encounters with such UXO has begun to rise.

Interest in exploring various sonar approaches to detecting and identifying UXO in such environments has been growing. Among the many issues now being studied at a number of laboratories are the following two fundamental questions: What are the broadband acoustic scattering characteristics associated with typical submerged UXO, and how are these impacted by the bottom sediment? This information is critical to the development and evolution of sonar systems able to detect submerged UXO on and in the sediment, and able to efficiently separate these detections from those due to natural and man-made clutter whose number density is expected to be fairly high in many underwater environments of interest.

Recently, Bucaro *et al.*[1] reported laboratory grade underwater acoustic scattering measurements on four UXO targets. These measurements, the first reported broadband, multiaspect, high precision echo measurements for submerged UXO, were made in the free-field, i.e., away from any boundary such as a sediment bottom. Further, the measurements were carried out *monostatically*, meaning that the source and receiver were co-located as the target echo returns were collected over a complete 360°.

In the present study, further echo measurements are obtained on two of these UXO targets, both to extend the measurements to *bistatic* echo responses, i.e., echoes detected by a receiver not co-located with the exciting source, and to begin to access the effect of the sediment on the target echo levels as well. Regarding the latter, we focus on incidence angles well below the critical angle for the water-sediment interface, a case relevant to long range sonars. Future studies will consider the above-critical angles relevant to shorter range, down-looking systems.

II. EXPERIMENTAL DETAILS

The measurements were conducted in NRL's 250 000 gal sediment laboratory pool facility, which has been described by Simpson *et al.*[2] The sediment facility has physical boundaries of $8 \times 10 \times 7$ m^3 deep, with a 3.8 m de-ionized water column over a sandy bottom 3 m deep that is filtered, washed, and well-characterized. The mean grain diameter is 240 μm, and the interface with the water can be leveled and smoothed to less than 0.5 mm rms roughness. With reference to Fig. 1, measurements were made with the target proud of the bottom for two source angles, 0° and 90°, as measured from the normal to the target's front end (left side of the targets in the photos).

Bistatic measurements of target scattering were made with the facility in its compact scattering range mode, as shown in the top diagram of Fig. 1. In this mode, broadband impulses are used to excite the source, and acoustic data are collected over a limited time window chosen to exclude energy returning from the walls of the pool facility. The extent of this window, which varied for different receiver/target an-

[a]Author to whom correspondence should be addressed. Electronic mail: bucaro@pa.nrl.navy.mil

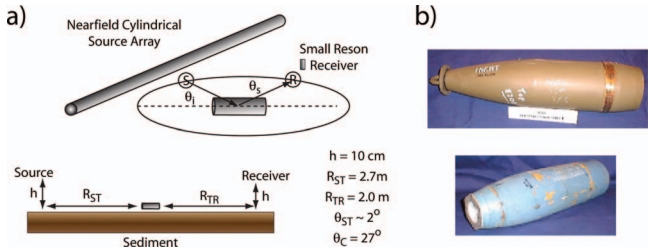


FIG. 1. (a) The experimental configuration for the bistatic measurements. The source angle θ_i and the scattering angle θ_s are measured from a normal to the target front end. The source is a horizontal near-field array, and the receiver is a small Reson hydrophone. (b) Photos of the 155 mm shell (top) and the 5 in. rocket (bottom) with the front of both targets on the left in the photos.

gular sectors, generally ranged from 1.5 to 2 ms. The target was placed proud of the sediment at the center of the facility, together with the source and receiver positioned 2.7 and 2.0 m, respectively, from the target both 10 cm above the sediment surface. The source is a horizontally mounted 3 m long near-field line array in which the transducer elements are shaded in such a way as to produce a pseudo-plane-wave sound field in the near-field of the array throughout a limited volume, centered at the target position. The line source generates a pseudo-unipolar broadband pressure pulse approximately 20 μs in duration, covering a band of frequencies in excess of 25 kHz. The receiver is a Reson 4013 small hydrophone attached to a vertical graphite composite pole mounted to an in-air rotation stage that is digitally positioned with a computer and stepper motor. The scattered echo response was measured 2.0 m from the target for a particular source direction θ_i as a function of scattering angle θ_s by rotating the receiver in 1° increments over 360°.

The data were collected and processed to recover complex scattering cross section expressible as TS referenced to 1 m. Three quantities are measured—the incident acoustic pressure, the pool clutter (background) pressure, and the scattered pressure—and these are measured in the following way. First, before positioning the target, the source is excited, and the incident pressure is measured at the location corresponding to the target center for the scattering measurement. Second, the source is excited, and the non-target pressure field is measured as a function of θ at each receiver position to be used in the scattering experiment. This measurement contains scattering from pool clutter and, in the forward scattering plane, the incident pressure field. Lastly, the target is inserted, and the scattered pressure field is measured as a function of scattering angle θ_s.

The target strength is obtained by first subtracting the non-target measurement from the scattering measurement. This process, together with imposition of the aforementioned time window, eliminates energy from (1) any indirect paths such as reflections from the finite-sized pool walls (including those which first reflect from the target), (2) submerged equipment, and (3) the direct incident wave for bistatic angles in the forward directions. This step is possible only through precise control of the locations of the acoustic elements, and only if fluctuations in the acoustic medium are sufficiently small. For our facility, robotic control of the source and receiver position is approximately 30 μm, and the iso-velocity water is maintained to within 0.01 °C for more than a 24 h period. With the non-target data file removed from the scattered signal, the parameter $X(f,\theta)$ is formed in terms of the scattered signal $P_{\text{scat}}(f,\theta)$ and the incident field measured at the target center $P_{\text{inc}}(f)$ as follows:

$$X(f,\theta) = \frac{P_{\text{scat}}(f,\theta)}{P_{\text{inc}}(f)} \frac{r_{\text{scat}}}{e^{ikr_{\text{scat}}}}, \tag{1}$$

where f is frequency, and r_{scat} is the distance from the target center to the receiver. The scattering data are measured at a range (2 m), which is in the near-field for some target aspects, and in the far-field for others. Since bistatic data can be readily projected to the far-field, we performed this projection on all the echo measurements. The target strength values are then defined and displayed as $10\log_{10}(|X(f,\theta)|^2)$.

The two UXO targets, which were also used in the monostatic measurements reported by Bucaro *et al.*,[1] are shown in the right portion of Fig. 1. One (upper) is a 155 mm artillery shell whose length and maximum diameter are 63 and 15.5 cm, respectively. The second (lower) is a 5 in. rocket warhead whose length and maximum diameter are 45 and 12.7 cm, respectively. Both are filled with a polymer material whose density, bulk modulus, and Poisson ratio are 1500 kg/m^3, 8.1×10^9 Pa, and 0.34, respectively. This material extends over the entire length of the shells except for an air cavity at the front, which is 13 and 5 cm deep in the 5 in. rocket and 155 mm projectile, respectively. Both targets have relatively thick steel walls (~1.4 cm), which become thicker at the back end.

III. RESULTS

Shown in Fig. 2 are the bistatic target strength values for each target displayed as $10\log_{10}(|X(f,\theta)|^2)$ in the proud case for the two source incidence angles, 0° and 90°.

A. 0° incidence and backscattering

For the 0° incidence case, we see two major highlights: the backscattered response at $\theta_s=0°$, which in the case of the 155 mm shell has significant levels above about 10 kHz, and the high response centered at $\theta_s=\pm180°$, i.e., the forward scattered direction.

As we discuss below, the first major highlight has components related to specular reflection from the more or less flat front of the targets, as well as contributions from other effects such as acoustic and/or elastic waves reflecting from the far end after traveling down the length of the target. This identification is supported by the frequency-time display shown in Fig. 3. In this display, the first component observed has the familiar strong support in time characteristic of specular reflection, followed by a number of other specular-like components delayed various amounts of time from the front-end specular. As can be seen in Fig. 3, in the case of the 155 mm shell, the front-end specular is by far the dominant mechanism, in contrast to that for the 5 in. rocket where all the components are of comparable level.

We can consider bistatic scattering from the target front end for more general θ_i using a simple model. For a flat-

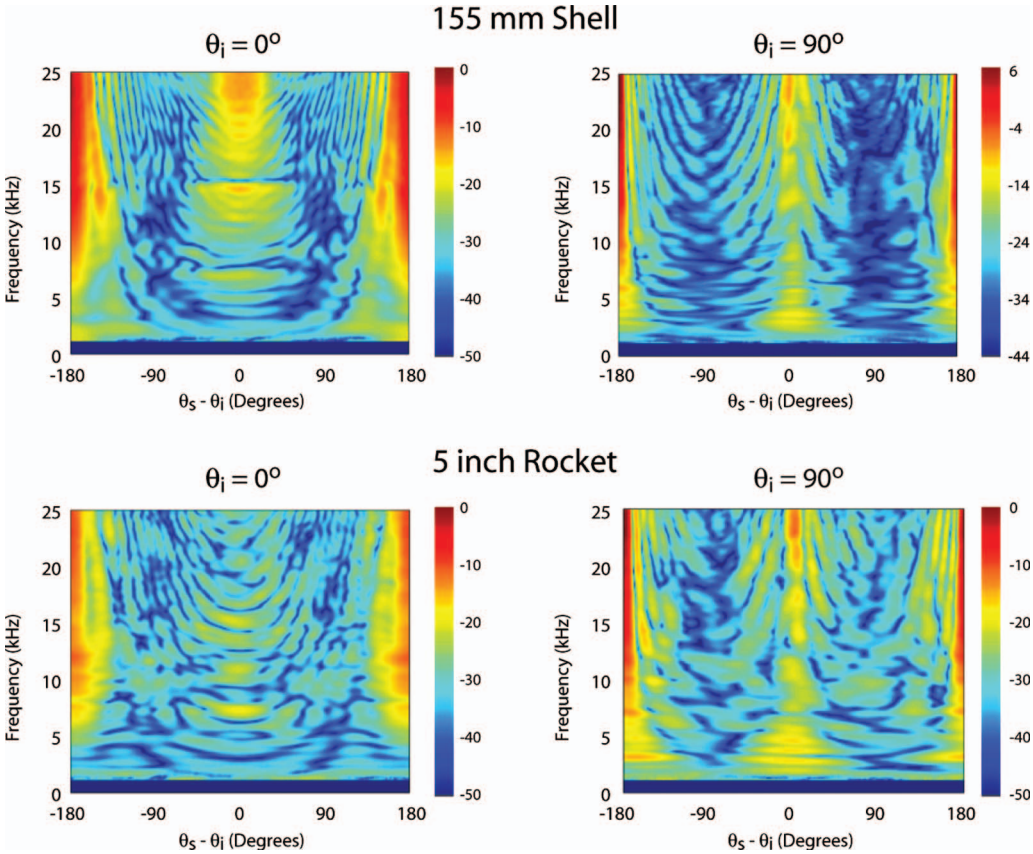


FIG. 2. Measured target strength for the proud targets color coded in dB versus frequency and scattering angle, minus source angle for θ_i=0° (left) and θ_i =90° (right) for the 155 mm shell (upper) and the 5 in. rocket (lower).

ended UXO insonified at θ_i, one can approximate the bistatic TS contribution from the ends in the backscattered plane for the special case of specular reflection (θ_s=360°−θ_i) and source angles 0° ≥ θ_i < 90°. This is done using expressions for the monostatic scattering from a rigid, circular disk modified by the projected area factor cos θ (θ being the angle between the incident or scattered wave and the normal to the disk). Using the monostatic reflection expression for the disk [Eq. 3a of Ref. 1 from Urick[3]], but now inserting the additional factor cos θ, gives

$$\mathrm{TS} = 20 \log \frac{A \cos \theta}{\lambda}, \tag{2}$$

where A is the area of the circular disk and λ is the acoustic wavelength.

For backscattering at normal incidence where θ=0, Eq. (2) predicts that the backscattered echo from the 155 mm shell front (assumed to be a 4 in. diameter circular disk) would rise with frequency at a level of 6 dB per octave, and Eq. (2) with θ=0 is plotted in Fig. 4, along with the measured data. As can be seen, the general agreement is good. The two pairs of peaks and nulls at ~7 and 14 kHz are most likely due to resonances of the disk-like front end. In fact, we calculate 7.9 and 16 kHz for the first two modes of vibration of a circular plate clamped at its edges having the same dimensions.[4] The slight lowering of the observed frequencies is probably due to the fluid and radiation loading ignored in this estimate.

We also observe a low level frequency modulation across the band. This is related to the other time delayed low level impulses seen in the frequency-time plots of Fig. 3. We believe that these are associated with elastic waves and creeping,[5] or diffracted waves traveling at their corresponding speeds down the length of the target and then reflecting back from the discontinuity provided by the back end. The time delay between the returning waves and the initial specular reflection from the front end would be twice the target length/wave speed, where the relevant speeds would be ~5941, 3251, 2991, 2824, 1390, and 1500 m/s. The first three are the longitudinal, shear, and surface wave speeds in the steel wall, the fourth and fifth are the longitudinal and shear wave speeds in the filler, and the last is the speed of sound in water for the creeping wave. In Fig. 3, for the 155 mm shell, one can find a time delayed signal close to each of these speeds, but attempting a definitive identification with the wave types is not warranted, given both the overall uncertainties and the possibility of mode conversion.

Next, we consider the 5 in. rocket. The details of the 155 mm shell differ in several ways from those of the 5 in. rocket, and we believe some of these differences lead to the different character discussed below in their echo responses. In particular, for the 5 in. rocket, the front disk-like area is smaller by a factor of ~2, and thinner by a factor 2.5. It does not stand off from the body as does the disk in the 155 mm

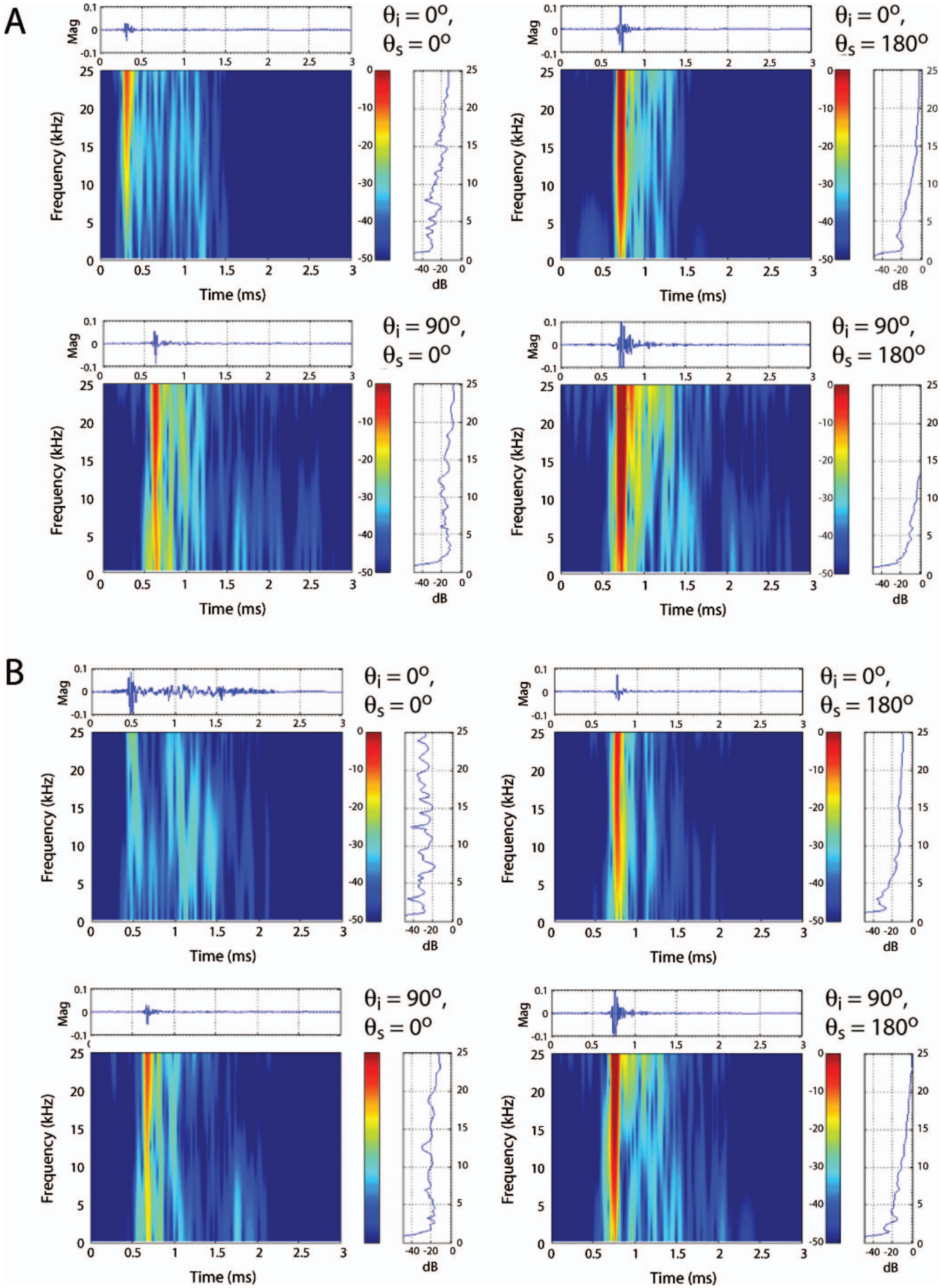


FIG. 3. Frequency-time transforms of the measured target strength for θ_i=0° (upper) and θ_i=90° (lower), and θ_s=0° (left) and θ_s=180° (right), for the 155 mm shell (a) and the 5 in. rocket (b).

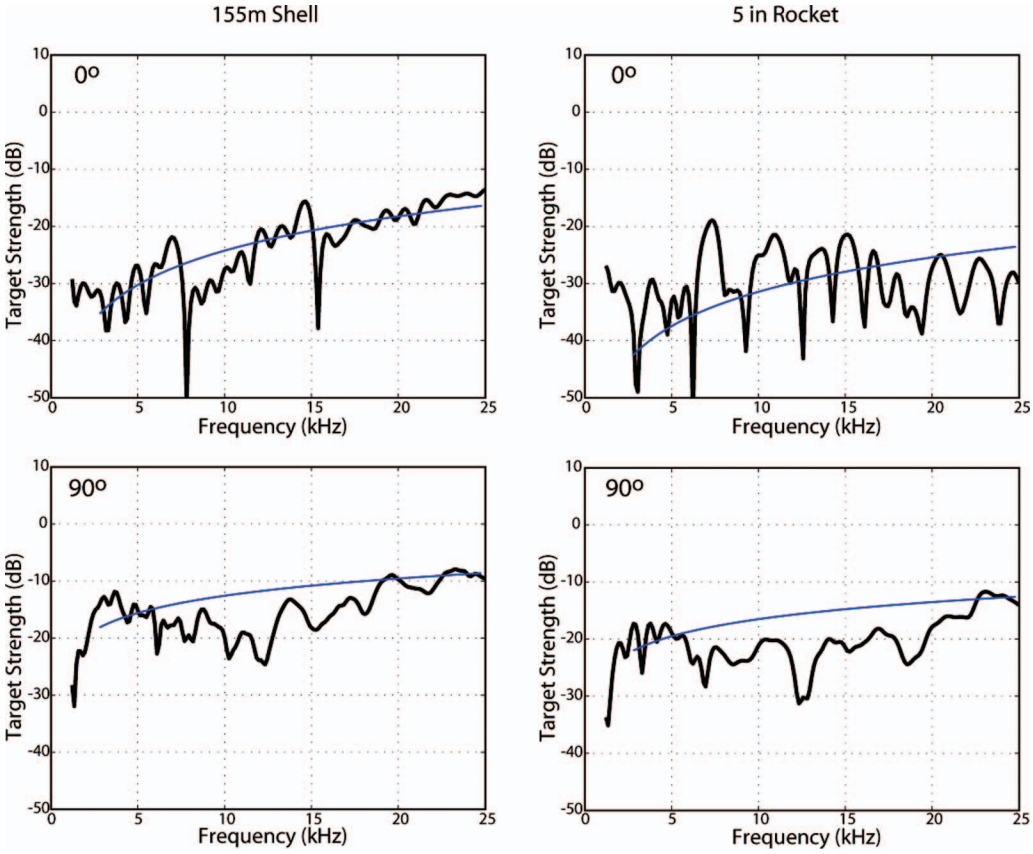


FIG. 4. Backscattered target strength in dB versus frequency for 0° incidence (upper) and 90° incidence (lower) for the 155 mm shell (left) and the 5 in. rocket (right). Lines are calculated using Eq. (2) (upper) and Eq. (4) (lower).

shell, and immediately following is a much deeper (~5 in.) air-filled cavity. Finally, the shell body is shorter (43.2 cm versus 63.5 cm) and much less tapered.

The backscattered echo at 0° for the 5 in. rocket is shown versus frequency in the upper right hand quadrant of Fig. 4, along with the prediction based on Eq. (2). As can be seen, the TS levels are reduced from that seen in the 155 mm shell by about 6 dB congruent with the front-end area ratio of 2 for these targets. Further, although the overall level agrees generally with the prediction, the oscillations with frequency now are large compared to those in the 155 mm shell case because of the reduced level of the front-end specular reflection. Indeed, unlike the specular return, which increases with frequency at 6 dB per octave [see the 155 mm return for the 0° case in Fig. 4 and the Eq. (2) prediction], the creeping wave[5] and elastic wave return peaks are expected to be more or less independent of frequency over this band as is actually observed for the overall level for the 5 in. rocket. As can be seen in the frequency-time plots of Fig. 3, the first after-specular response of comparable level arrives about 0.6 ms later, a time consistent with both the creeping wave speed and the filler shear speed. We can attempt to differentiate between the contributions of the two components in the returns by using the TS computed numerically for a rigid version of this target reported by Bucaro *et al.*[1] The rigid case for 0° allows only specular, and the creeping wave scattering, and we find that above about 15 kHz, the rigid calculation has levels and oscillations similar to those seen in Fig. 4. We therefore conclude that at these frequencies, the echo is principally due to specular scattering and creeping waves.

The origin of the final strong return seen in the frequency-time plots of Fig. 3, about 1 ms after the specular, is unknown. We find that all elastic waves in the steel wall, as well as the longitudinal wave in the filler, would arrive much sooner and, thus, cannot be the source of this late return. We considered the possibility of a return of a slow airborne wave from the back of the 13 cm long air-filled cavity just behind the front in this target. However, the associated time delay would be ~1.5 ms, which is too long. In addition, the level of the return, being comparable to the specular, would require a disk/cavity transmission factor ~1, which is not realizable at the metal/air interface of this target. Perhaps this late return is related to mode conversions of the various elastic waves.

The large deviations from the specular predictions discussed above (the two resonances for the 155 mm shell and the large fluctuations with frequency for the 5 in. rocket) are also observed qualitatively in the free-field monostatic measurements at 0° in Ref. 1, supporting the conclusion that the effects are not due to the presence of the sediment but to the details of the target structure.

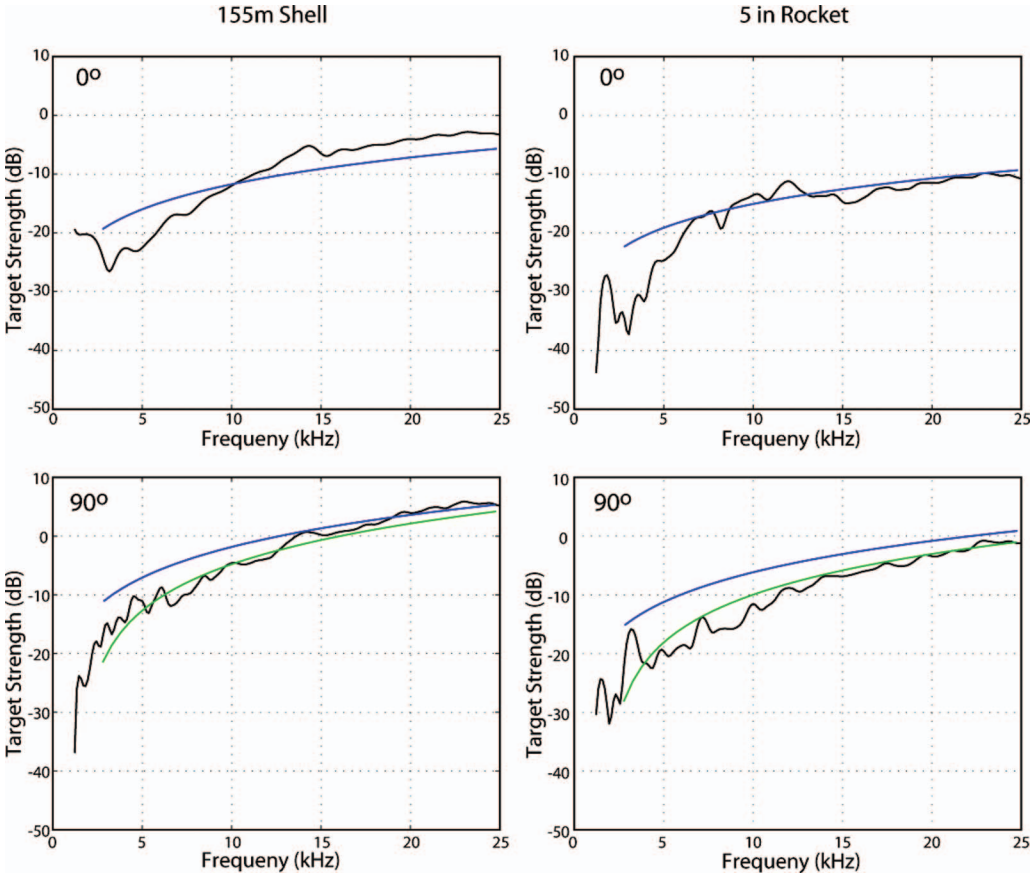


FIG. 5. Forward scattered target strength in dB versus frequency for 0° incidence (upper) and 90° incidence (lower) for the 155 mm shell (left) and the 5 in. rocket (right). Lines are calculated using Eq. (3) (upper) and Eq. (5) (lower), assuming soft (blue) or rigid (green) boundary conditions.

B. 0° incidence and forward scattering

The second and highest component in the 0° incidence case is seen at $\theta_s = \pm 180°$, i.e., for the forward scattered component. We can predict what one would expect for the forward scattering of an acoustic signal incident on the end of the cylinder using the equivalent analysis of Ross[6] as applied to the electromagnetic (radar) scattering from a perfectly conducting cylinder.[7] For this case, the forward scattered TS depends on the target's projected area under the curve separating the illuminated and shadowed regions (as predicted from simple physical optics) with a correction term, which takes into account the curved cylindrical surface, which is also illuminated by the incident wave. Unlike the beam incidence case, which we will discuss later, symmetry here eliminates dependence on polarization in the radar case. For the acoustic case, this implies that for end-on incidence, acoustically soft and rigid boundary conditions lead to identical results.[7] For 0°, Ross' result can be shown for our acoustic case to give

$$\mathrm{TS} = 10 \log \left| \sqrt{\sigma^{PO}(0°)} \exp\left(\frac{j\pi}{2}\right) + \sqrt{\sigma^{CS}(0°)} \exp\left(\frac{j3\pi}{4}\right) \right|^2, \tag{3}$$

where $\sigma^{PO}(0°) = \pi^2 (a^2/\lambda)^2$, $\sigma^{CS}(0°) = 0.3(\pi a L)(a/\lambda)$, and a and L are the cylinder radius and length. The first term is the physical optics result, and the second term takes into account the contribution from the curved surface of the cylinder. For the dimensions of the 5 in. rocket, and approximating its shape as cylindrical, Eq. (3) gives TS=−9 dB at 25 kHz, which compares well with the measured value of −10 dB, especially given the fact that the UXO is not a perfect cylinder. Likewise, using Eq. (3) for the somewhat larger 155 mm shell predicts a forward scattered TS at 25 kHz of −5.5 dB, also in reasonable agreement with the measured value of −3.5 dB.

We also show the frequency dependence of the exactly forward scattered signal in the upper portion of Fig. 5 along with the predictions of Eq. (3) for both targets. As can be seen, above about 10 kHz, the agreement is reasonably good, especially for the 5 in. rocket. In this band, for either target, the data and the prediction have roughly the same frequency dependence. At low frequencies, however, the data fall off more rapidly than the prediction as one lowers frequency. Below 5 kHz, the second term in Eq. (3) (due to scattering from the cylindrical surface, which decreases only 3 dB per octave as frequency is lowered) dominates, becoming twice that of the physical optics term (which decreases by 6 dB per octave). We believe that the difference between the analytic expression and measurement is related to overestimation of the cylindrical surface scattering effect, which in turn may be related to the taper of these targets.

In the physical optics approximation, the forward scattered target strength depends directly (and only) on the projected area, and the phase is purely imaginary. The target

strength is then simply 20 log(area/λ). At 25 kHz, this approximation predicts forward scattering levels of −13.5 and −10.0 dB for the 5 in. rocket and 155 mm shell, respectively. Both of these values are several decibels lower than the total forward scattered TS observed, indicating that scattering along the length of the cylinder does contribute a non-trivial amount to TS, compared to the larger physical optics occulted area effect. Indeed, the first and second terms in Eq. (3) predict that at 25 kHz, side-wall scattering is as much as ~80% of that associated with the physical optics effect for both the 5 in. rocket and 155 mm shell.

C. 90° incidence and backscattering

Next, we consider the 90° source angle case. As can be seen here again, there are only two dominant highlights: the backscattered highlight from the "beam" of the cylinder and the stronger forward scattered component.

To predict the bistatic scattering from the cylindrical section of a rigid (or soft) cylinder over the range θ_i and $\theta_s = (0, \pi)$, one can use the expression given by Junger,[8] viz.,

$$\mathrm{TS} = 20 \log\left\{ \frac{L}{2} \sin\theta_i \left[\frac{2ka}{\pi} (\sin\theta_i + \sin\theta_s)^{-1} \right]^{1/2} j_0\left[\frac{kL}{2} (\cos\theta_i + \cos\theta_s) \right] \right\}, \quad (4)$$

where θ_i and θ_s are again the source and receiver angles, respectively, and L is the cylinder length.

The frequency dependence of the 90° backscattered TS is shown in Fig. 4 (lower plots), along with the predictions of Eq. (4) for the two targets. In using Eq. (4), we have taken a as the maximum radius of the target, and L as the total length minus the lengths of the strongly tapered sections. As can be seen, for either target, the predictions and the measurements agree at the low and high ends of the band. However, both results exhibit a broad depression in between, as well as added fluctuations with frequency.

With respect to the latter, the fluctuations with frequency are almost certainly related to elastic effects in these two thick-walled shells. For example, circumnavigating elastic waves with speed C would produce a frequency modulation given by $C/2\pi a$, the reciprocal of the time to travel one circumference. Thus, circumferential shear and longitudinal waves excited in the wall would produce backscattered components with spectral periodicities of about 7 and 12 kHz for the 155 mm shell, and 8 and 15 kHz for the 5 in. rocket. For either target, the modulation produced by the flexural wave would go from 3 to 10 kHz from the lowest to highest frequency.

Regarding the former, the mid-frequency depression may be related to interference between scattering from the direct incident signal and that from the incident energy, which first reflects at a shallow angle off the sediment prior to striking the cylinder-like targets. For the geometry and associated distances used in our measurements, we estimate a difference of about 0.07 m for these two paths. This is equal to λ/2 and λ at 11 and 21 kHz, respectively, and, thus, such an effect is consistent with the overall frequency dependence of the depression. One might expect a similar effect for the end-on incidence case, and the fact that this is not apparent may be related to the planar versus cylindrical surface presented.

D. 90° incidence and forward scattering

For 90° incidence, Ross'[6] expression for forward scattering cross section of an electromagnetic field (radar) with wavenumber k would give

$$\mathrm{TS} = 10 \log\left| \sqrt{\sigma^{PO}(90°)} W \exp\left(\frac{j\pi}{2}\right) + \sqrt{\sigma^{\mathrm{side}}} \exp\left(\frac{j3\pi}{4}\right) \right|^2, \quad (5)$$

where $\sigma^{PO}(90°) = 4(aL/\lambda)^2$, $W = W_{hh} \sim 1 + 0.498(ka)^{-2/3} - 0.011(ka)^{-4/3}$ with $\sigma^{\mathrm{side}} = 0$ for horizontal polarization, and $W = W_{vv} \sim 1 - 0.432(ka)^{-2/3} - 0.214(ka)^{-4/3}$ with $\sigma^{\mathrm{side}} = 7/\pi(a^3/\lambda)$ for vertical polarization. As described by Bowman *et al.*,[7] for an infinitely long, perfectly conducting cylinder, the electromagnetic result for horizontal or vertical polarization is equivalent to the acoustic result for a soft or rigid cylinder, respectively, and we take this as applicable to our finite cylinder. The unity term in the expression for W represents the physical optics result, while the terms in inverse powers of ka take into account the effect of the curved surface near the shadow boundary.

We show in Fig. 5 the frequency dependence of the beam scattered forward scattering TS, as measured and as predicted from Eq. (5) for both soft and rigid cases. One can see that the data agree with the prediction for the rigid case fairly well above about 4 kHz (i.e., above $ka \sim 1$), and given the relatively thick steel walls of these targets, this is not surprising. It is important to point out that while the effects of elastic waves (shear, flexural, circumferential, etc., waves in the target casing) appear to some degree in beam backscattering, the imprint of such elastic waves on the forward scattered spectrum would be reduced by an order of magnitude because of the corresponding increase in the geometrically forward scattered signal. Clearly, these shells are neither perfectly rigid nor perfectly soft. However, given our two forward scattering models, the data are seen to fall closer to the rigid predictions.

Unlike the axial incidence case, we point out that the physical optics approximation, based simply on occulted area, now gives a prediction close to the more exact expression in Eq. (5). In particular, at 25 kHz, for example, the simpler physical optics result gives −0.3 and 4.3 dB for the 5 in. rocket and 155 mm shell, respectively, compared to the −1.3 and 3.2 dB values from Eq. (5).

Regarding these proud target measurements, we make the following two observations, which can be taken as general only to the extent that we have captured the salient behavior of the bistatic mechanisms by sampling just two source angles (0° and 90°). First, given the rough agreement between the predicted backscattering levels, which do not take into account the sediment (especially recognizing the deviation of these targets from a perfect cylinder), for the

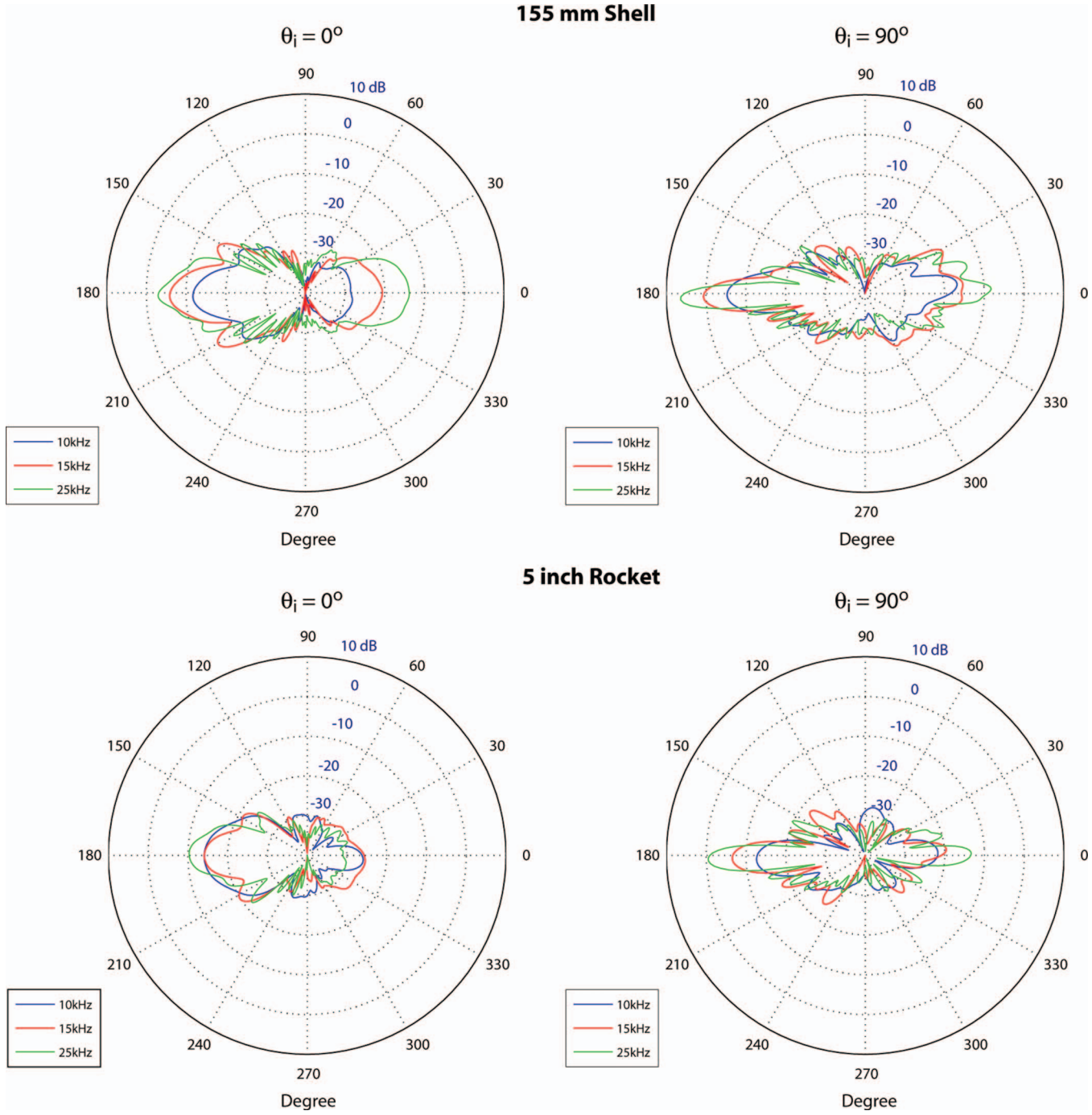


FIG. 6. Bistatic beam patterns for 0° and 90° incidences at three frequencies for both targets.

proud targets, the effect of the sediment on TS does not appear to be large. This is consistent with other results we have obtained, comparing free-field and proud cases for monostatic measurements.[9] Further, in general, the sediment effects tend to *increase* TS, although not always. Second, over most of the band, we see the largest target strength by far in the bistatic case, specifically for the forward scattered echo. This can be seen even more clearly in the plots in Fig. 6, showing the bistatic scattering beam patterns at three frequencies for the complete 360°.

IV. DISCUSSION AND CONCLUSIONS

Our motivation for collecting and considering the data presented above was to be able to infer if there is merit in exploiting bistatic target scattering vis-à-vis monostatics for the purpose of detecting submerged proud UXO, and to determine how the broadband acoustic scattering characteristics associated with submerged UXO might be impacted by the presence of the sediment.

In addressing the sediment issues, it is important to distinguish between two conditions as related to the sediment-water interface, viz., whether one is operating above or below the critical angle $\theta_C = \cos^{-1}(C_f/C_s)$, where C_f and C_s are the sound speeds in the fluid and in the sediment, respectively. For the sediment in the facility used in these measurements $\theta_C = 28.1°$, and given our range and height above the sediment, the associated measurements have been made well

below θ_C. For this condition, sound penetrates into the sediment only evanescently and only to a depth of about one wavelength.

The crude estimates for detection range shown in Fig. 10 of Ref. 1 can be shown to predict that in our measurement band, a TS ≥ 0 dB leads to signal to noise (S/N) ratios ≥ 20 (an approximate criterion compatible with target detection) at 100 m ranges in a typical noise environment with a modest source level (170 dB re: μPa). For the proud UXO's studied here, only one scattering component approaches these 0 dB TS levels, viz., the broadband forward scattered beam echo.

Another look at Fig. 2, together with Eqs. (2)–(5), suggests the following conclusions regarding the exploitation of bistatic target scattering: (1) In general, the bistatic response in the backscattered plane appears to provide at least comparable levels of broadband target strength relative to the monostatic scenario reported in Ref. 1. Accordingly, the bistatic response for targets proud on the sediment should provide target strength levels with sufficient S/N levels for detection and identification at reasonable ranges. In this regard, for specular scattering from the ends of the target, Eq. (2) shows that as one moves away from normal incidence (0°) toward 90°, echo levels, which are proportional to cos θ_i, fall off weakly at first. At the same time, Eq. (4) shows that specular scattering from the target side, which is proportional to $\sin^{1/2}\theta_i$, falls off even more weakly with angle away from normal incidence. (2) By far, the largest bistatic responses are seen at forward angles (180° from the source direction). They are as much as 10 dB higher than any monostatic return. Although we have detailed expressions (and measurements) at forward scattering for only 0° and 90° incidence, we point out that the major term for the latter across the band and for the former except at the lowest frequencies is the physical optics term given by the projected area of the target. For intermediate source angles, we can therefore expect the forward scattered TS to remain relatively high, and to fall somewhere between the values observed at 0° and 90°.

We conclude from this limited study that access to bistatic echo information in operations aimed at detecting submerged UXO targets could provide an important capability. The measurements carried out here in the laboratory used a source and receiver just 10 cm above the sediment, about 2 m from the target, giving an associated grazing angle of ~2°, which is well below the critical angle for our sediment (27°). For the longer ranges typical of an actual field measurement, these distances would scale so that the same grazing angle would be achieved, for example, for a target range of 30 m by a source and receiver positioned 1.5 m above the sediment, a height compatible with practical sonar systems.

The largest target strength levels are associated with the forward scattered echo. However, obtaining an exploitable forward scattered echo is considerably more difficult than doing so for backscattered signals because of the simultaneous presence of the strong incident field. We overcame this difficulty in our laboratory measurements in a straightforward manner because we were able to map the details of the incident field with high precision *before* placing the target into position. Clearly, this is not possible in a search for UXO targets in the environment. However, the high forward scatter to backscattered target strength ratios, which persist in the presence of the sediment, and perhaps even increase upon burial, warrant serious consideration of how this might be accomplished. Field approaches, which attempt to extract the incident field in the forward direction, include, for example, mode filtering in a water channel[10] and apex shifted Radon transforms,[11] as applied to ground penetrating radar. The former requires long vertical arrays, which are not practical for our application. The latter capitalizes on the different hyperbolic range-cross range dependences of the source and scattered signals. We are currently exploring techniques related to the latter.

ACKNOWLEDGMENTS

This work was carried out under support from the Strategic Environmental Research and Development Program (SERDP) and ONR. J.A.B. is an on-site contractor for EXCET, Inc. (Springfield, VA 22151). L.K. and T.Y. are on-site contractors for Global Strategies Group.

[1] J. A. Bucaro, B. H. Houston, M. Saniga, L. R. Dragonette, T. Yoder, S. Dey, L. Kraus, and L. Carin, "Broadband acoustic scattering measurements of underwater unexploded ordnance (UXO)," J. Acoust. Soc. Am. **123**, 738–746 (2008).
[2] H. J. Simpson, B. H. Houston, and R. Lim, "Laboratory measurements of sound scattering from a buried sphere above and below the critical angle," J. Acoust. Soc. Am. **113**, 39–42 (2003).
[3] R. Urick, *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York, 1983), pp. 17–30.
[4] W. Soedel, *Vibrations of Shells and Plates* (Dekker, New York, 1981), pp. 103.
[5] I. V. Andronov and D. Bouche, "On the degeneration of creeping waves in a vicinity of critical values of the impedance," Wave Motion **45**, 400–411 (2008).
[6] R. A. Ross, "Forward scattering from a finite, circular cylinder," Electromagn. Waves **2**, 207–215 (2008).
[7] J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi, *Electromagnetic and Acoustic Scattering by Simple Shapes, Revised Printing* (Hemisphere, New York, 1987), pp. 89–91.
[8] M. C. Junger, "Formulation of short-wavelength bistatically scattered fields in terms of monostatic returns," J. Acoust. Soc. Am. **95**, 3055–3058 (1994).
[9] J. A. Bucaro and B. H. Houston, "Sonar systems for prosecuting underwater UXO," in Partners in Environmental Technology Technical Symposium and Workshop, Washington, DC (2008).
[10] A. Sarkissian, "Extraction of a target scattering response from measurements made over long ranges in shallow water," J. Acoust. Soc. Am. **102**, 825–832 (1997).
[11] F. Yong, Z. Zheng-ou, and X. Jia-li, "Clutter reduction based on apex shifted radon transform in sub-surface forward-looking ground penetrating radar," in International Conference on Radar CIE'06 (2006), pp. 1–3.

Acoustic emission source location in composite structure by Voronoi construction using geodesic curve evolution

R. Gangadharan, G. Prasanna, M. R. Bhat, C. R. L. Murthy, and S. Gopalakrishnan
Department of Aerospace Engineering, Indian Institute of Science, Bangalore 560012, India

(Received 30 March 2009; revised 15 August 2009; accepted 17 August 2009)

Conventional analytical/numerical methods employing triangulation technique are suitable for locating acoustic emission (AE) source in a planar structure without structural discontinuities. But these methods cannot be extended to structures with complicated geometry, and, also, the problem gets compounded if the material of the structure is anisotropic warranting complex analytical velocity models. A geodesic approach using Voronoi construction is proposed in this work to locate the AE source in a composite structure. The approach is based on the fact that the wave takes minimum energy path to travel from the source to any other point in the connected domain. The geodesics are computed on the meshed surface of the structure using graph theory based on Dijkstra's algorithm. By propagating the waves in reverse virtually from these sensors along the geodesic path and by locating the first intersection point of these waves, one can get the AE source location. In this work, the geodesic approach is shown more suitable for a practicable source location solution in a composite structure with arbitrary surface containing finite discontinuities. Experiments have been conducted on composite plate specimens of simple and complex geometry to validate this method. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3224736]

PACS number(s): 43.40.Le, 43.40.Sk, 43.40.Dx, 43.60.Jn [EJS] Pages: 2324–2330

I. INTRODUCTION

Acoustic emission (AE) technique is a prevalent nondestructive testing tool for dynamic assessment of component's health condition. The specific merits such as global monitoring capability and passive nature of sensing make it a preferable technique for real-time monitoring.[1] But less work has gone into making this technique as an quantifying tool. As it is well evaluated and proven that AE signals are reliable indicators of physical change in the component under study, better results and understanding can be expected if the geometrical location of the signal is also ascertained with accuracy. It helps to predict more accurately the physical phenomenon the system is undergoing and its significance.[2–5] The present work is focused on obtaining the location of AE source in a composite structure.

One of the major advantages of AE technique as an online monitoring tool is its capability to locate active defects in larger structural components without having to physically scan them.[6] Different methods based on analytical, numerical, and empirical approaches have been proposed to locate discrete AE events in the structure.[7] Tobias[8] used the triangulation method to locate the AE source. The same approach was extended to a spherical surface by Asty.[9] A mathematical formulation based on the concept of geodesics was applied to a cylindrical surface for AE source location.[10] An alternative location technique uses the concept of "the first sensor hit by an AE event" to identify a more generalized region around each sensor, from which the AE signal likely originated. In this case, one can determine which one of the several sensor regions on the test specimen has more concentrated AE activity.[11] Some AE systems determine signal arrival times using fixed threshold techniques[12] and because of the aforementioned complications, such AE systems measure arrival times for signals using various portions of the AE signal, which travel at different velocities. These approaches are affected by the signal attenuation and dispersion of elastic waves due to inhomogeneity and geometry of the material.[13] Cross-correlation techniques were used to obtain the arrival-time difference of the dispersive AE modes for source location.[14] Time-frequency analysis techniques such as wavelet transform were used to study the dispersion characteristics of AE modes in order to compute the group speed of the propagating wave modes accurately. This information was used along with triangulation method for AE source location in thin metallic plates.[15] The same approach was employed for AE source location in a thin composite plate using only two AE sensors.[16] Soft computing methods such as artificial neural networks have been used for source location, source identification, and defect severity classification.[17] An alternative approach based on an optimization scheme[18] was proposed to locate the point of impact in isotropic and anisotropic plates. Almost all the triangulation based approaches require analytical/parametric representation of the surface geometry for the formulation of governing equations. They provide little options in terms of geometrical variation and hence limited to simple geometry. Even when parametric representation is available, it is practically time-consuming and error-prone to solve either analytically or numerically the governing equations of complex geometries such as aircraft structures. Additionally, none of the above approach can be applied to geometries with discontinuities such as sharp corners, holes, etc. (which are very common in most structures).

In the present work, circular piezoelectric sensors are arranged as a spatially distributed array in the structure. A graph-theory based concept has been employed here to compute the discrete geodesic path using Dijkstra's algorithm in

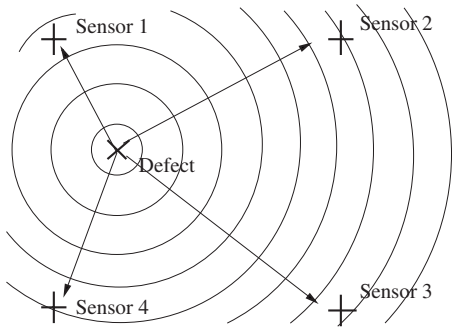


FIG. 1. Wavefront propagation from defect location.

a triangulated meshed surface of the structure. Geodesic is defined as the shortest path between two points in the space. In the plane, the geodesics are straight lines and on the sphere, the geodesics are great circles. The Voronoi diagram concept is then used to determine the AE source location by allowing the geodesics to propagate from each sensor location until the source is reached. The Voronoi diagram for a set of points S in the plane is a partition of the plane into convex polygons, each of which consists of all the points in the plane closer to one particular point of S than to any other.[19] This study aims to locate the AE source accurately in an anisotropic thin composite structure of planar and complex geometries.

II. APPROACH

The approach suggested here is based on the fact that the wave takes minimum energy path to travel from a source to any other point in the connected domain. This conceptual view can be visualized as in Fig. 1, which shows waves propagating in all possible directions from a defect location, along the minimum energy path. With imposed material limitations, this path is generated by geodesics. Once the geodesic paths are extracted in a given geometry, the defect location is reached by back-propagating the waves along those paths from the sensor locations. To start with, it can be assumed that sensors detect only surface waves, which in turn means that the study is restricted to two-dimensional (2D) surfaces. It is to be noted that even the triangulation method utilizes the same approach, only that it has an inherent assumption that the geometry is a three-dimensional (3D) continuum resulting in simple distance equation based on Pythagorean theorem, which is solved analytically or numerically. Since our approach builds on the fact that a wave takes minimum energy path, the shortest energy path reduces to shortest-distance path, which is provided by evaluating the geodesics. Mathematically, the energy along a path is seen as weighing function $w(x)$ defined along the path. Hence, the minimum energy path is given by the following.

For energy along path,

$$L = \int w(x)dx. \tag{1}$$

For minimum energy path,

$$\min(L) = \min\left(\int w(x)dx\right), \tag{2}$$

where $\min(\int w(x)dx)$ is the required *geodesic*. So, the minimum-energy-path problem is equivalent to shortest-distance-path problem. For anisotropic solids the weighting function is direction dependent. The weighting function can be a function of parameters such as direction dependent stiffness, group velocity, or slowness profile of the propagating Lamb wave modes. Though there definitely exists a mapping between stiffness and slowness, characterization of a material distribution through kinematic behavior (slowness curve) is more amenable to experimental determination compared to that of employing kinetic behavior (stiffness). Furthermore, the degradation of the elastic properties with time due to introduction of damage, heat, etc., makes it difficult to compute the direction dependent stiffness of the composite structure. In the present work, the slowness profile of the Lamb wave mode is used for computing the weighting function.[20] The weighting function can be defined linear over the triangular element and with mesh refinement better results can be obtained. Quadratic or higher order weighing function can definitely be employed over coarser mesh to get similar results. Since the objective of this paper is primarily to convey the concept of employing weighting function for handling anisotropies, we have limited the discussion to linear weighting function.

Here we consider a two stage strategy of computing the geodesics in the structure using graph theory followed by locating the AE source through the intersection of geodesics from the sensor array based on Voronoi diagram concept.[21] The strategy for geodesic computation and Voronoi diagram concept is explained in Sec. II A and II B.

A. Formulation-wave propagation approach

This approach involves discretizing the domain as curved or planar simplicial complex chains,[19] followed by finding local geodesics in each of the simplex and finally gluing them together to get the required global geodesic. The approach can be viewed as a wave emanating at a point, searching for the shortest path in the neighborhood and moving to it, continuing this until it reaches the destination. There are many suggested methods to calculate discrete geodesics and it is still an active area of research investigation.[23–25] In this work, the simpler Dijkstra's algorithm[22] has been used to compute the discrete geodesics in the discretized structure.

Consider two sensors $S1$ and $S2$ located at (S_{1x}, S_{1y}) and (S_{2x}, S_{2y}) with V being the velocity of wave in the medium. If a source S is generated at location at (S_x, S_y) in a planar surface then the distance between the source and sensor $S1$ is given by

$$D(S1 - S) = \sqrt{(S_{1x} - S_x)^2 + (S_{1y} - S_y)^2}, \tag{3}$$

and the distance from the source to sensor $S2$ is given by

$$D(S2 - S) = \sqrt{(S_{2x} - S_x)^2 + (S_{2y} - S_y)^2}. \tag{4}$$

Let t_1 and t_2 be the time taken by the wave generated from the source to reach the sensors $S1$ and $S2$, respectively. The difference in time for the wave reaching the two sensors given by $dt = t_1 - t_2$ is proportional to distance difference be-

tween the source and the sensors. Hence, Eqs. (4) and (5) are combined to get

$$D(S1-S)-D(S2-S)=Vdt. \tag{5}$$

The only unknowns in Eq. (5) are the locations of the source. Hence by forming enough distance-time-difference equations we can solve for the source location. The number of difference equations (N) that can be formed and number of sensors (n) used are related by $N=n(n-1)/2$. For example, in 3D source location, we have (x,y,z) as the unknowns to be determined. Using two sensors will give $N=2\times(2-1)/2=1$ equation(s). Hence at least three sensors are to be used, which gives $N=3\times(3-1)/2=3$ equations. Mostly a numerical approach, such as least-squares, is used to solve the equations and in addition more number of sensors than the required minimum helps to get an improved positional accuracy. This can also be seen in GPS where four or more satellites are employed to calculate a geographical location, against the required minimum of three. It is important to observe that the distance D in above formulation is nothing but the geodesic distance. It can be written as an implicit function given by

$$\Phi(D,V,dt)=0. \tag{6}$$

Thus Eq. (6) combines the information from geometry of the surface (geodesics) and the information from the material (velocity), with the experimental observation (dt). The geodesics D in above formulation is arrived from wave-propagation perspective using graph-theory based Dijkstra's algorithm. The governing equation is subsequently further recast as

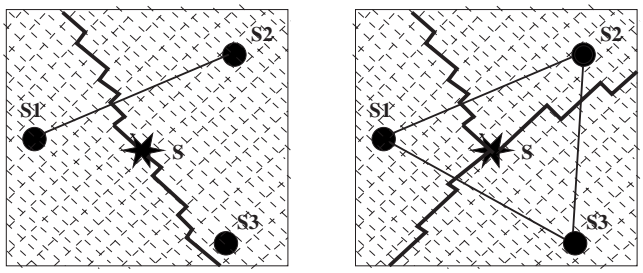
$$D(S1-S)\pm Vdt=D(S2-S), \tag{7}$$

leading to the view that the solution lies in the boundary of the Voronoi diagram and the exact location is at the intersection of two more boundaries. The next part is the construction of Voronoi like diagram to locate the intersection of wavefronts, which is discussed in Sec. II B.

B. Voronoi construction

The formulation derived in Sec. II A is applied to locate all points that are equidistant from the sensor locations resulting in a Voronoi diagram.[19] A simple search for the node having the least error in distance from all the sensors is the intersection point and chosen as the location of the source. The following workout elucidates the Voronoi construction approach.

Taking a case of three-sensor setup provides three sets of time-difference equations. Using the recast formulation, all the points that are equidistant from sensor locations are found (using distance map calculated based on Dijkstra's algorithm along with Vdt corrections). When information from only two sensors is available, then only one equation is formulated and hence there exist multiple solutions meeting the distance criteria. This is depicted in Fig. 2 with the jagged thick line passing between sensors $S1$ and $S2$ indicating all points that are equidistant (which is the Voronoi diagram). An important observation is that this line passes through the



Two sensors Three sensors

FIG. 2. Voronoi diagram.

defect location (AE source) and hence we need to search only along this line for getting to the source. When information from one more sensor ($S3$) is considered, then three equations are formulated, from which the other lines in Voronoi diagram are constructed. All these lines intersect exactly at the AE source, as in Fig. 2. The geodesic lines joining the three sensors give the Delaunay triangle, which is the dual of conventional Voronoi diagram. Above construction is trivial to implement in form of set operations. For a given mesh, we arrive at the following.

(1) Let nk be the kth node in the mesh.
(2) Let $D(nkSi)$ be the distance between kth node and ith sensor.
(3) Let $Dk(S_{ij})$ be the difference in distances of a node k from sensors Si and Sj, i.e., $Dk(S_{ij})=D(nkSi)-D(nkSj)$.
(4) Voronoi line between any two sensors Si and Sj is formed by nodes that satisfy the condition that $Dk(S_{ij})=Vdt_{ij}$ where dt_{ij} is the hit arrival-time difference between the sensors Si and Sj and D's are geodesic distances and the corresponding line can be seen as set of these nodes, which is given by

$$L_{ij}=\{nk|Dk(S_{ij})=Vdt_{ij}\}. \tag{8}$$

(5) Hence for three sensors we get L_{12}, L_{13}, and L_{23} as shown in Fig. 2. The intersection point is the intersection point in the set L_{ij} given by

$$\text{source},\quad S=\{n|(L_{12}\cap L_{23}\cap L_{13})\}. \tag{9}$$

(6) For surfaces, which are intrinsically 2D in parametric space, only two of the above sets are to be included for getting the source node.

The geodesic approach discussed above is validated by performing experiments on composite plate specimens, and the accuracy of the algorithm in locating the AE source location is studied.

III. EXPERIMENTAL STUDY

Experiments were conducted on thin composite plates of simple and complex geometry. When the thickness of the plate becomes comparable to the wavelength of propagating waves, guided waves such as Lamb waves are sustained in these plates. The schematic of the experimental setup shown in Fig. 3 consisted of lead zirconate titanate (PZT) wafers acting as transmitter and receivers, preamplifiers to amplify the signal from the sensors, and a NI-PXI 6115 data acqui-

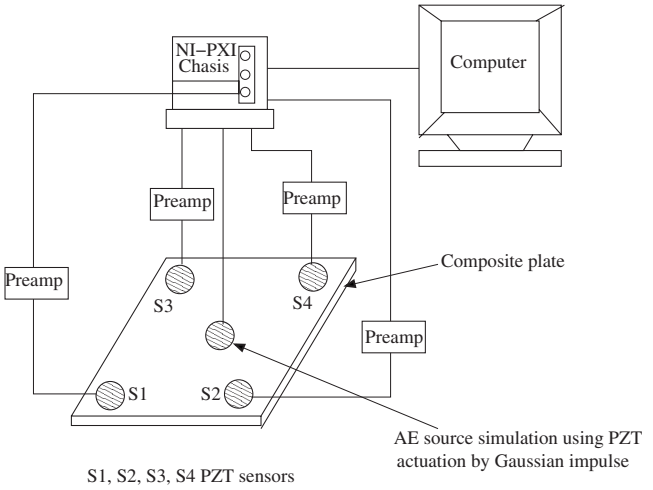


FIG. 3. Schematic of the experimental setup.

sition card. Using the analog output channel of the NI-DAQ card, a tone burst signal was generated to actuate the PZT wafers. The signal picked up by the receiver PZT sensor was amplified and transferred to the computer via the NI-DAQ card. The PZT wafer sensors used were 10 mm in diameter and 1 mm in thickness. The sensors were surface bonded to the structure using phenyl-salicylate salt. For the algorithm to work for anisotropic composite materials, one has to obtain the group velocity values in all the direction. A compact array of single transmitter and multireceiver (STMR) (Ref. 20) system has been used to measure the Lamb wave group velocity in different directions. A circular array of 24 PZT sensors and diameter 150 mm equispaced at 15° were surface bonded on the plate. From the experimental signals, extraction of A_0 mode was difficult as the amplitude of the signal was very small. Hence, only S_0 mode propagation time-of-flights were measured at 200 kHz and used to calculate group velocities of S_0 mode in different directions. In this work, the AE source was simulated in the structure by applying a broadband Gaussian impulse signal to the PZT wafer surface bonded to the structure. The simulated source generated predominantly the S_0 mode in the plate and the velocity, and time information of the S_0 was used for geodesic computations. Furthermore, we assumed that no mode conversion of Lamb wave modes takes place in the structure.

A. Glass/epoxy composite plate

Experiments were conducted on a quasi-isotropic composite plate $[0/90/45/-45/0/90]_S$ of dimensions $240\times240\times2.4$ mm^3. The group velocity profile of the above composite plate obtained using STMR system is shown in Fig. 4. The weighting function used for computing the geodesic is same for all directions as the variation of group velocity is constant. The PZT sensors $S1$, $S2$, and $S3$ were bonded at locations ($x1=0$ mm, $y1=240$ mm), ($x2=240$ mm, $y2=240$ mm), and ($x3=240$ mm, $y3=0$ mm), respectively, on the plate. The AE source was simulated at the position ($x0=80$ mm, $y0=80$ mm) and the signals acquired by the sensors are shown in Fig. 5. The first noticeable peak of the S_0 mode above the noise level is considered as the arrival

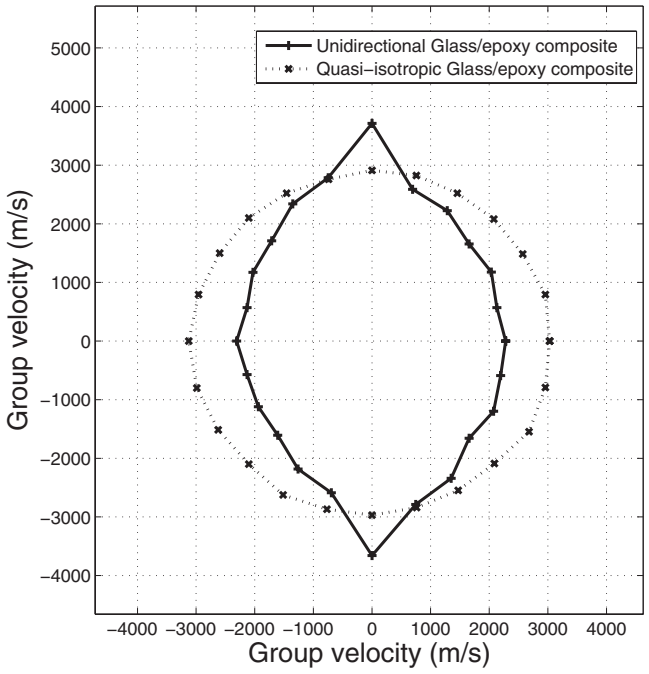


FIG. 4. Group velocity profile for glass/epoxy composite plate.

time for the signals considered above. The arrival-time difference information between the sensors (dt_{12}, dt_{13}, and dt_{23}) is obtained from the time history of the signals acquired by the sensors. The plate was meshed with triangular elements using ANSYS finite element software and the mesh information was used to obtain the geodesics. Dijkstra's algorithm was then applied to compute the shortest paths from the sensors. To begin with a coarse mesh was used for computation of the discrete geodesic path. In this method, the wave propagation is considered to take place only along the edges of the triangle. The location of the AE source is shown in Fig. 6(a) and shows the minimum path taken from the sensors to the AE source. The shortest path for the coarse mesh did not come out to be straight lines between AE source and sensors $S1$ and $S3$. In order to come closer to the actual path, the triangular elements in the mesh were refined. The geodesic path obtained for finer mesh [Fig. 6(b)] was closer to the actual path when compared to the coarse mesh results. The

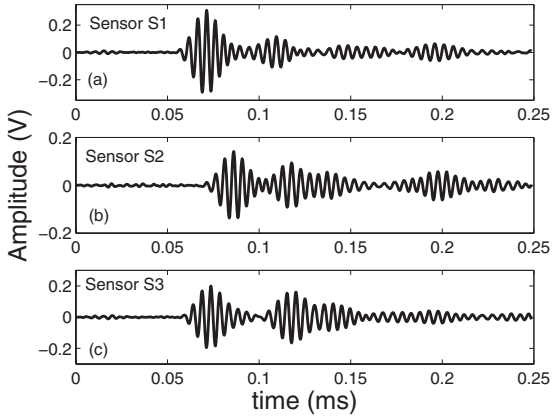


FIG. 5. Quasi-isotropic glass/epoxy composite plate: Signals acquired by the (a) sensor $S1$, (b) sensor $S2$, and (c) sensor $S3$.

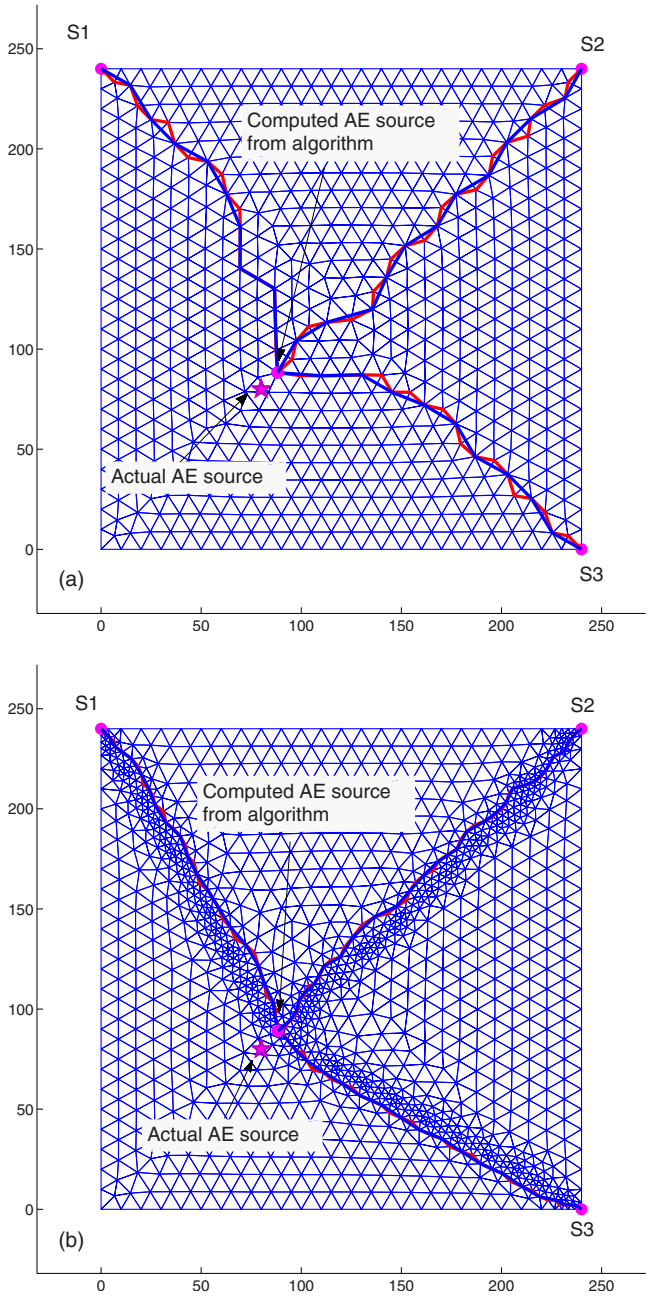


FIG. 6. (Color online) Quasi-isotropic glass/epoxy composite plate. Source location by geodesic propagation: (a) coarse mesh and (b) fine mesh.

source location computed by the algorithm is given by (x =88.33 mm, y=89.18 mm) for coarse mesh and (x =85.21 mm, y=85 mm) for fine mesh, and the results are closer to the actual source location ($x0$=80 mm, $y0$ =80 mm). As a result of refinement there was improvement in source location results and also convergence to the actual geodesic path between the AE source and sensors.

Experiments were then conducted on a uni-directional composite plate $[0]_{12}$ of dimensions $380\times350\times2.4$ mm^3. The group velocity profile of the composite plate obtained using STMR system is shown in Fig. 4. The slowness curve is then evaluated to define the weights in all directions. For uni-directional composite, the weight is minimum along the fiber direction as the group velocity is high compared to other directions. The PZT sensors $S1$, $S2$, $S3$, and $S4$ were

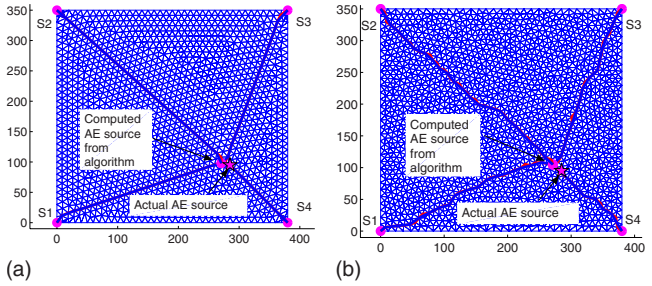


FIG. 7. (Color online) Uni-directional glass/epoxy composite plate. Source location by geodesic propagation: (a) mesh configuration 1 (MATLAB) and (b) mesh configuration 2 (ANSYS).

bonded at locations ($x1$=0 mm, $y1$=0 mm), ($x2$ =350 mm, $y2$=0 mm), ($x3$=380 mm, $y3$=350 mm), and ($x4$=380 mm, $y4$=0 mm), respectively, on the plate. The AE source was simulated at the location ($x0$ =285 mm, $y0$=95 mm). In order to show the effect of the mesh sensitivity in the computation of geodesics, two different mesh configurations were considered. The mesh configurations were generated using MATLAB and ANSYS softwares. The shortest path taken by the wave from the sensors and their intersection is shown in Fig. 7. There was not much difference in the AE source location for both the mesh configurations. However, the paths taken by the waves to reach the source were different as they have the limitation of propagating along the edges, which depends on the mesh configuration. The experimental results are summarized in Table I.

B. Bi-directional woven carbon/epoxy composite plate with a hole

Experiments were conducted on a bi-directional woven carbon/epoxy composite plate of dimensions 280×280 $\times2.5$ mm^3. The group velocity profile of the composite plate obtained using STMR system and based on the velocity profile of the weighting function is computed. A hole of diameter 25 mm was introduced in the center of the plate. The PZT sensors $S1$, $S2$, $S3$, and $S4$ were bonded at locations ($x1$=0 mm, $y1$=0 mm), ($x2$=140 mm, $y2$=120 mm), ($x3$=0 mm, $y3$=280 mm), and ($x4$=280 mm, $y4$ =280 mm), respectively, on the plate. The AE source was simulated at the location ($x0$=140 mm, $y0$=220 mm) such that the sensor $S2$ was placed behind the hole in the plate. The plate was meshed and the mesh information was used to obtain the geodesics. The shortest path taken by the wave from the sensors and their intersection are shown in Fig. 8.

TABLE I. Source location by geodesic algorithm-glass/epoxy composite.

	Source location		Algorithm result		
Cases	$x0$ (mm)	$y0$ (mm)	X (mm)	Y (mm)	Error (%)
Quasi-isotropic (coarse mesh)	80	80	88.33	89.18	10.95
Quasi-isotropic (fine mesh)	80	80	85.21	85	6.38
Uni-directional mesh 1	285	95	270.85	98	4.12
Uni-directional mesh 2	285	95	271.28	104.83	3.19

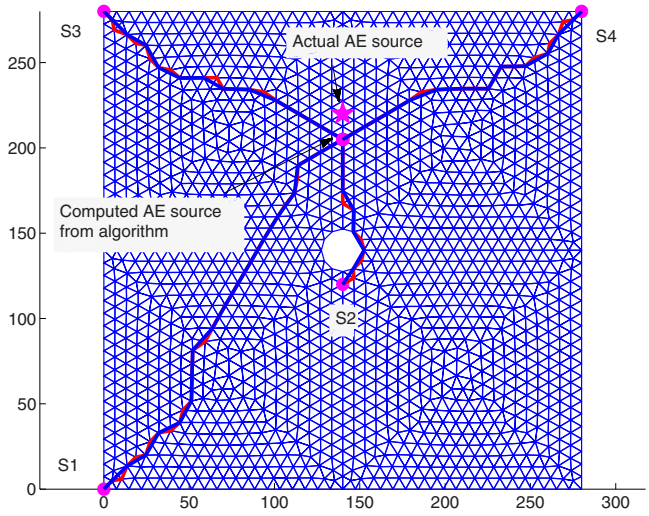


FIG. 8. (Color online) Bi-directional woven carbon/epoxy composite plate: AE source location by geodesic propagation.

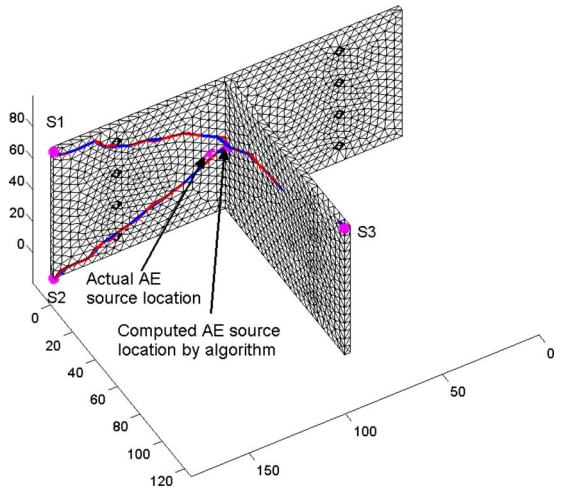


FIG. 10. (Color online) T-pull carbon/epoxy composite specimen-3D: Source location by geodesic propagation.

The geodesic path from the sensor $S2$ clearly shows that the wave traveled around the defect to reach the source. This result shows the capability of the algorithm to handle surfaces/objects with structural discontinuities. The source location computed by the algorithm is given by (x =140 mm, y=220 mm), and the result shows a deviation of 3.59% from the actual position ($x0$=140 mm, $y0$ =205 mm). However, if the hole scatters the wave completely, the sensor $S2$ fails to pick up any signal resulting in the sensor being placed in the shadow region. But using the information from the other sensors, the AE source location can be evaluated.

C. T-pull composite specimen

Experiments were conducted on a T-pull carbon/epoxy composite specimen. The PZT sensors $S1$, $S2$, and $S3$ were bonded at locations ($x1$=190 mm, $y1$=0 mm, $z1$ =3 mm), ($x2$=190 mm, $y2$=80 mm, $z2$=3 mm), and ($x3$=90 mm, $y3$=0 mm, $z3$=110 mm), respectively, on the specimen. The AE source was simulated at the location ($x0$=100 mm, $y0$=40 mm, $z0$=3 mm). The group velocity profile of S_0 mode was obtained experimentally, and the variation of group velocity was found to be 2500 ± 100 m/s. So a constant group velocity of 2500 m/s was assumed in all directions. Also, we ignore the variation in group velocity and mode conversion phenomenon occurring in the central region of the T-pull specimen where there is a change in thickness. The T-pull specimen was meshed with 2D triangular elements, and the mesh information was used to obtain the geodesics. The shortest path taken by the wave from the sensors and their intersection are shown in Fig. 9. The source location computed by the algorithm is given by (x =90 mm, y=37.5 mm, z=4.31 mm), and the result shows a deviation of 9.39% from the actual position ($x0$ =100 mm, $y0$=40 mm, $z0$=0 mm).

The algorithm was then extended to 3D case. The T-pull specimen was meshed using 3D tetrahedral elements, and the geodesics were computed on them. The geodesic path from the sensors and their intersection are shown in Fig. 10. The source location computed by the algorithm is given by (x =91.5 mm, y=42.5 mm, z=7.22 mm), and the result shows an error of 6.12%. The assumption of constant velocity in all directions and limitation of the wave to travel along the edges can be attributed to the presence of error in the results. The experimental results for the carbon/epoxy composite specimens are listed in Table II.

From the above experimental results the geodesic method can be easily extendable to planar surfaces as the computation of weighting function from the slowness profile is simple. But for specimens with complicated geometry it is

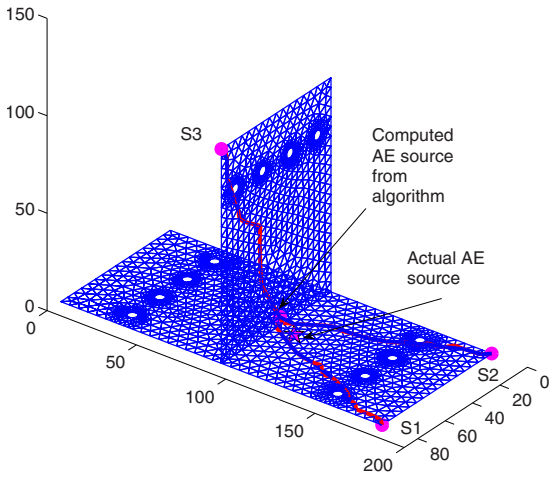


FIG. 9. (Color online) T-pull carbon/epoxy composite specimen-2D: Source location by geodesic propagation.

TABLE II. Source location by geodesic algorithm-carbon/epoxy composite specimen.

Cases	Source location $x0$ (mm)	$y0$ (mm)	$z0$ (mm)	Algorithm result x (mm)	y (mm)	z (mm)	Error (%)
Bi-directional woven plate	140	220	0	140	205	0	4.8
T-pull specimen (2D)	1001	40	0	90	37.5	4.31	9.39
T-pull specimen (3D)	100	40	3	91.5	42.5	7.22	6.12

difficult to define the weighting function using the slowness profile or direction dependent stiffness. Hence certain assumptions are made to get an approximate weight function, which affects the accuracy of the source location. Therefore the difficulty in obtaining the exact weighting function limits the application of geodesic method to complex anisotropic structures.

IV. CONCLUSION

The approach presented in this study used the geodesic property of wave and the Voronoi diagram concept to locate the AE source in composite structures. Graph theory was used to compute the discrete geodesics on the triangular mesh using Dijkstra's algorithm. The relation of geodesics to source location problem was established by proving that location of source as the first intersection point of multiple geodesics generated from the sensors location using the Voronoi diagram concept. The algorithm was experimentally validated on planar and odd geometry composite specimens. The convergence of the discrete geodesics to the actual geodesic path between the sensors and the AE source with mesh refinement was demonstrated. The location of an AE source with less error in the plate with a hole problem shows the ability of the method to handle surfaces/objects with structural discontinuities. The approach was then extended to an anisotropic T-pull specimen and the geodesic method was studied using both 2D and 3D triangular mesh elements. The results obtained by the method were closer to the actual source location, and the presence of error is attributed to the assumptions involved in computing the weighting function and the limitation of the wave to travel along the edges of the triangular mesh. The accuracy of the algorithm can be further improved by extending the geodesic wave to travel over the triangulated domain instead of moving along the edges of the triangle. It can be asserted that the geodesic curve-evolution method holds great promise for versatile implementation catering to non-conventional geometries.

ACKNOWLEDGMENT

The work is carried out under the project supported by the Boeing Co.

[1]R. Geng, "Modern acoustic emission technique and its application in aviation industry," Ultrasonics **44**, e1025–e1029 (2006).
[2]M. Giordano, A. Calabro, C. Esposito, A. D'Amore, and L. Nicolais, "An acoustic-emission characterization of the failure modes in polymer composite materials," Compos. Sci. Technol. **58**, 1923–1928 (1998).
[3]Y. H. Yu, J. H. Choi, J. H. Kweon, and D. H. Kim, "A study on the failure detection of composite materials using an acoustic emission," Compos. Struct. **75**, 163–169 (2006).
[4]A. Bussiba, M. Kupiec, S. Ifergane, R. Piat, and T. Bohlke, "Damage evolution and fracture events sequence in various composites by acoustic emission technique," Compos. Sci. Technol. **68**, 1144–1155 (2008).
[5]M. A. Hamstad, "A review: Acoustic emission, a tool for composite-material studies," Exp. Mech. **26**, 7–13 (1986).
[6]S. McBride, Y. Hong, and M. Pollard, "Enhanced fatigue crack detection in aging aircraft using continuous acoustic emission monitoring," Rev. Prog. Quant. Nondestr. Eval. **12**, 2191–2197 (1993).
[7]P. J. Shull, *Nondestructive Evaluation: Theory, Techniques and Applications* (Dekker, New York, 2002).
[8]A. Tobias, "Acoustic emission source location in two dimensions by an array of three sensors," Nondestr. Test. **9**, 9–12 (1976).
[9]M. Asty, "Acoustic emission source location on a spherical or plane surface," NDT Int. **11**, 223–226 (1978).
[10]P. Barat, P. Kalyanasundaram, and B. Raj, "Acoustic emission source location on a cylindrical surface," NDT & E Int. **26**, 295–297 (1993).
[11]M. A. Hamstad and K. S. Downs, "On characterization and location of AE sources in real size composite structures—A waveform study," J. Acoust. Emiss. **13**, 31–41 (1995).
[12]B. Castagnede, W. Sachse, and K. Y. Kim, "Location of pointlike AE sources in anisotropic plates," J. Acoust. Soc. Am. **86**, 1161–1171 (1989).
[13]M. Ohtsu, "AE theory for moment tensor analysis," Res. Nondestruct. Eval. **6**, 169–184 (1995).
[14]S. M. Ziola and M. R. Gorman, "Source location in thin plates using cross-correlation," J. Acoust. Soc. Am. **90**, 2551–2556 (1991).
[15]H. Jeong and Y. S. Jang, "Fracture source location in thin plates using the wavelet transform of dispersive waves," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 612–619 (2000).
[16]N. Toyama, J. H. Koo, and R. Oishi, "Two-dimensional AE source location with two sensors in thin CFRP plates," J. Mater. Sci. Lett. **20**, 1823–1825 (2001).
[17]P. P. Delsanto, *Universality of Nonclassical Nonlinearity: Application to Non-Destructive Evaluations and Ultrasonics* (Springer, New York, 2006).
[18]T. Kundu and S. Das, "Point of impact prediction in isotropic and anisotropic plates from the acoustic emission data," J. Acoust. Soc. Am. **122**, 2057–2066 (2007).
[19]J. A. Sethian, *Level Set Methods and Fast Marching Methods* (Cambridge University Press, Cambridge, MA, 2002).
[20]J. Rajagopalan, K. Balasubramaniam, and C. V. Krishnamurthy, "A phase reconstruction algorithm for Lamb wave based structural health monitoring of anisotropic multilayered composite plates," J. Acoust. Soc. Am. **119**, 872–878 (2006).
[21]G. Prasanna, M. R. Bhat, and C. R. L. Murthy, "Acoustic emission source location on an arbitrary surface by geodesic curve evolution," J. Acoust. Emiss. **25**, 224–230 (2007).
[22]E. W. Dijkstra, "A note on two problems in connection with graphs," Numer. Math **1**, 269–271 (1959).
[23]K. Polthier and M. Schmies, "Straightest geodesics on polyhedral surfaces," International Conference on Computer Graphics and Interactive Techniques, ACM-SIGGRAPH Courses (2006), pp. 30–38.
[24]G. V. V. Ravi Kumar, P. Srinivasan, V. Devaraja Holla, K. G. Shastry, and B. G. Prakash, "Geodesic curve computations on surfaces," Computer Aided Graphics Design **20**, 119–133 (2003).
[25]V. Surazhsky, T. Surazhsky, D. Kirsanov, S. J. Gortler, and N. Hoppe, "Fast exact and approximate geodesics on meshes," ACM Trans. Graphics, **24**(3), 553–560 (2005).

Ultrasonic field modeling by distributed point source method for different transducer boundary conditions

Tamaki Yanagita and Tribikram Kundu[a)]
Department of Aerospace and Mechanical Engineering, University of Arizona, Tucson, Arizona 85721

Dominique Placko
SATIE, Ecole Normale Superieure, 61 Avenue Du President Wilson, F-94235 Cachan Cedex, France

(Received 9 December 2008; revised 6 July 2009; accepted 8 July 2009)

Several investigators have modeled ultrasonic fields in front of transducers by Huygens–Fresnel superposition principle that integrates the contributions of a number of point sources distributed on the transducer face. This integral solution, also known as the Rayleigh integral or Rayleigh–Sommerfeld Integral solution, assumes the strengths of the point sources distributed over the transducer face. A newly developed technique called distributed point source method (DPSM) offers an alternative approach for modeling ultrasonic fields. DPSM is capable of modeling the field for prescribed source strength distribution as well as for prescribed interface conditions with unknown source strengths. It is investigated how the ultrasonic field in front of the transducer varies in different situations: (1) when the point source strengths are known, (2) when the point source strengths are unknown but obtained from the interface condition that only the normal component of the transducer velocity is continuous across the fluid-solid interface, (3) when all three components of velocity are assumed to be continuous across the interface for the no-slip condition, and (4) when the pressure instead of the velocity is prescribed on the transducer face. Results for these different interface conditions are compared with the analytical solutions along the central axis.
 [DOI: 10.1121/1.3203307]

PACS number(s): 43.40.Rj, 43.20.Fn, 43.20.Bi, 43.20.El [TDM] Pages: 2331–2339

I. INTRODUCTION

Modeling of ultrasonic and sonic fields generated by a planar transducer of finite dimension is a fundamental problem whose solution is available in textbooks in this area.[1–4] A good review of the earlier developments of the ultrasonic field modeling in front of a planar transducer can be found in Ref. 5. A list of more recent developments is given by Sha *et al.*[6] The pressure field in front of a planar transducer in a homogeneous isotropic fluid has been computed both in the time domain[5,7,8] and in the frequency domain.[9–14] In the frequency domain analysis, which is more popular, the transducers are assumed to be vibrating with constant amplitudes at certain frequencies, and the pressure fields in front of the transducers are computed. Most of these models give the steady state response of standard transducers. Recently, the phased array transducers have been modeled as well.[15–20]

In all modeling approaches that are based on Huygens–Fresnel superposition principle and Rayleigh–Sommerfeld integral (RSI) solution, the strengths of the point sources modeling the transducer are assumed to be known. For type I integral the pressure distribution is assumed to be known, and for type II integral the velocity distribution is the known quantity. In distributed point source method (DPSM) analysis, the source strengths are assumed to be unknown and obtained from satisfying the interface conditions. Three different interface conditions are considered: (1) fluid and solid particle velocities at the interface are continuous in the perpendicular direction while slippage or discontinuity in the velocity components parallel to the interface is possible, (2) all three velocity components are continuous across the fluid-solid interface giving rise to no-slip condition, and (3) pressure at the transducer surface is specified, instead of velocity.

For an ideal fluid, it is not possible to satisfy the no-slippage interface condition across the fluid-solid interface. However, if the fluid has a small viscosity, then all three components of velocity must be continuous across the interface. What is the effect of this change in the interface condition on the computed ultrasonic field? Note that small viscosity of the fluid will give rise to the no-slippage condition. However, to keep the analysis simple, if one ignores the fluid viscosity during the field computation should one observe any noticeable change in the computed pressure field for changing the interface condition to no-slip condition? This question is answered in this paper.

Instead of the constant velocity, if the pressure is assumed to be uniform on the transducer face what will be its effect on the computed ultrasonic field in front of the transducer? These changes in the interface condition can be easily modeled by the DPSM (Refs. 21–25) and are carried out here.

If the cylindrical transducer element or the piezo-crystal has high stiffness, then it should contract and expand uniformly when excited by an electric pulse, resulting in uniform velocity at the transducer surface. For viscous fluids all

[a)]Author to whom correspondence should be addressed. Also at Department of Civil Engineering and Engineering Mechanics, University of Arizona, Tucson, Arizona 85721. Electronic mail: tkundu@email.arizona.edu

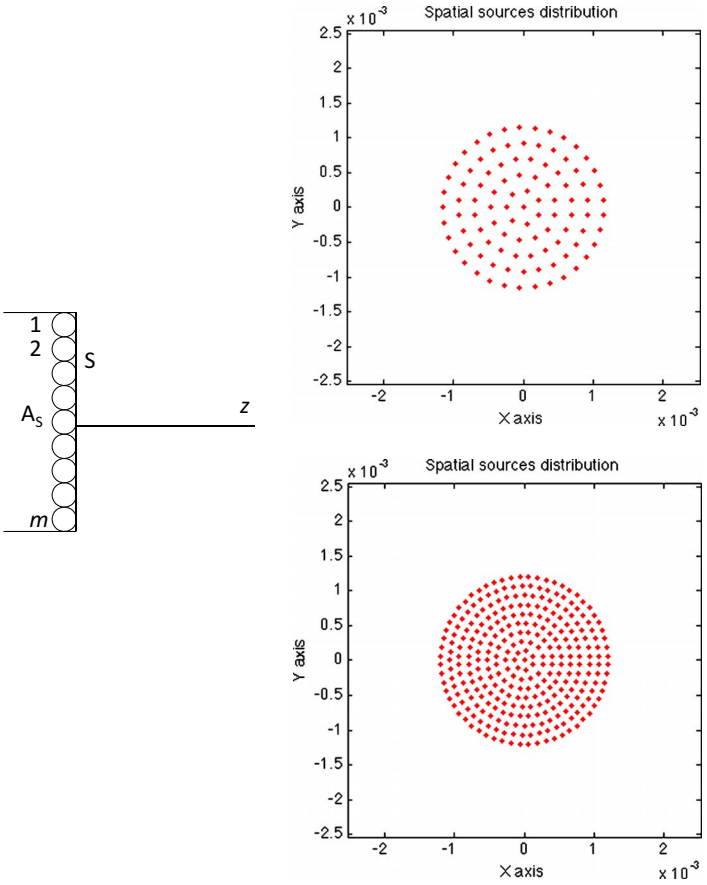


FIG. 1. (Color online) Left figure: In DPSM modeling, a finite transducer source S is modeled by a distribution of m point sources (the source layer is denoted as $\mathbf{A}_S$) placed behind the transducer face, as shown in the figure. Point sources are distributed on a plane slightly behind the front face of the transducer. Right figure: distribution of 100 (top) and 300 (bottom) point sources for modeling a circular transducer.

three components of velocity should be continuous across the fluid-solid interface, and for non-viscous fluids only the normal component is continuous, while slippage may occur for the parallel components across the interface. Therefore, a transducer vibrating in viscous and non-viscous fluids will excite the fluid particles adjacent to the transducer face differently.

Instead of a hard crystal, if a flexible disk (also known as the resilient disk in acoustic literature[26,27]) is used as the transducer then uniform pressure condition at the transducer face is generated.

II. THEORY

A solid transducer is placed in a fluid medium. To generate the acoustic field in the fluid medium, in front of the transducer face by the DPSM technique, a number of point sources are placed inside the solid transducer slightly behind the transducer face, as shown in Fig. 1. Point sources are located at the centers of the small spheres that touch the transducer face.[21–25] If the sphere radius is r_s, then clearly all point sources are placed at a distance, r_s, behind the transducer face. Such placement of point sources moves the singularities of the Green's functions outside the domain of interest, which is the fluid medium. Point sources are uniformly distributed over the total surface area of the transducer so that every point source approximately covers the same elemental area of the transducer surface. The first source is located at the center of the transducer face; then, a number of concentric circles are drawn with equal spacing to cover the entire surface area of the transducer. The point sources are then placed on the perimeters of these circles with the spacing between two neighboring point sources on the perimeters equal to the spacing between the concentric circles, as shown in Fig. 1. Point source locations for two cases are shown—when the transducer surface is modeled by 100 point sources and 300 sources. Ultrasonic field values (displacement, velocity, and pressure) at any point in the fluid medium can be obtained simply by superimposing the contributions of every point source. Velocity and pressure values obtained at the apex points of the spheres that touch the transducer face are equated to the prescribed velocity or pressure distributions to obtain the strength of individual point sources.

Following the derivation and notations presented by Placko and Kundu[21,22] in the DPSM formulation, the velocity ($\mathbf{V}_S$) and the pressure ($\mathbf{P}_S$) at the transducer surface can be computed from the point source strengths $\mathbf{A}_S$, as given below,

$$\mathbf{V}_S = [\mathbf{V}_{SX}, \mathbf{V}_{SY}, \mathbf{V}_{SZ}]^T = \mathbf{M}_{TSS}\mathbf{A}_S,$$

$$\mathbf{P}_S = \mathbf{Q}_{SS}\mathbf{A}_S. \tag{1}$$

The first subscript T of $\mathbf{M}_{TSS}$ stands for "total." Note that $\mathbf{M}_{TSS}$ gives the total (all three components of the) velocity

field. If one is interested in computing only one component of velocity ($\mathbf{V}_{SZ}$ for example), which is normal to the transducer surface, then from Eq. (1) only that component should be considered,

$$\mathbf{V}_{SZ} = \mathbf{M}_{SS}\mathbf{A}_S. \tag{2}$$

$\mathbf{M}_{SS}$ of Eq. (2) can be derived from $\mathbf{M}_{TSS}$ of Eq. (1) in the following manner:

$$\mathbf{M}_{TSS} = \begin{bmatrix} \mathbf{M}_{SSx} \\ \mathbf{M}_{SSy} \\ \mathbf{M}_{SSz} \end{bmatrix}_{3m\times m}, \quad \mathbf{M}_{SS} = [\mathbf{M}_{SSz}]_{m\times m}, \tag{3}$$

where m is the number of point sources used to model the transducer face or the source (see Fig. 1).

To obtain this relation for a source of complex shape whose normal direction varies from point to point, first a matrix is constructed by taking the direction cosines of the normal vectors on the source surface,

$$\mathbf{N} = [\text{DiagNorm}] = \lfloor \mathbf{n}_x, \mathbf{n}_y, \mathbf{n}_z \rfloor = \left[\begin{bmatrix} n_{x1} & 0 & 0 & \dots & 0 \\ 0 & n_{x2} & 0 & \dots & 0 \\ 0 & 0 & n_{x3} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & n_{xm} \end{bmatrix}, \begin{bmatrix} n_{y1} & 0 & 0 & \dots & 0 \\ 0 & n_{y2} & 0 & \dots & 0 \\ 0 & 0 & n_{y3} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & n_{ym} \end{bmatrix}, \begin{bmatrix} n_{z1} & 0 & 0 & \dots & 0 \\ 0 & n_{z2} & 0 & \dots & 0 \\ 0 & 0 & n_{z3} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & n_{zm} \end{bmatrix} \right], \tag{4}$$

where the jth point on the source surface has the normal direction (n_{xj}, n_{yj}, n_{zj}), where j varies from 1 to m if m point sources are used to model the source. Then,

$$\mathbf{M}_{SS} = \mathbf{N}\mathbf{M}_{TSS}. \tag{5}$$

Note that for the problem geometry shown in Fig. 1,

$$\mathbf{N} = \left[\begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \right]_{m\times 3m}. \tag{6}$$

Substituting $\mathbf{N}$ of Eq. (6) and $\mathbf{M}_{TSS}$ of Eq. (3) into Eq. (5), one obtains

$$\mathbf{M}_{SS} = \mathbf{N}\mathbf{M}_{TSS} = \left[\begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \right]_{m\times 3m} \times \begin{bmatrix} \mathbf{M}_{SSx} \\ \mathbf{M}_{SSy} \\ \mathbf{M}_{SSz} \end{bmatrix}_{3m\times m} = [\mathbf{M}_{SSz}]_{m\times m}. \tag{7}$$

If the velocity $\mathbf{V}_{SZ}$ at the transducer surface is specified as $\mathbf{V}_{S0}$ and the fluid adjacent to the transducer face is assumed to have the same velocity, then from Eq. (2),

$$\mathbf{M}_{SS}\mathbf{A}_S = \mathbf{V}_{S0}. \tag{8}$$

From Eq. (8),

$$\mathbf{A}_S = [\mathbf{M}_{SS}]^{-1}\mathbf{V}_{S0}. \tag{9}$$

Note that Eqs. (8) and (9) imply that the fluid particles adjacent to the transducer face have the same z-direction velocity as the transducer face, but there is no constraint on the fluid velocities in x and y directions.

If the fluid adjacent to the transducer is assumed to have the same velocity vector as the transducer face, in other words, if all three components of the velocity field are to be

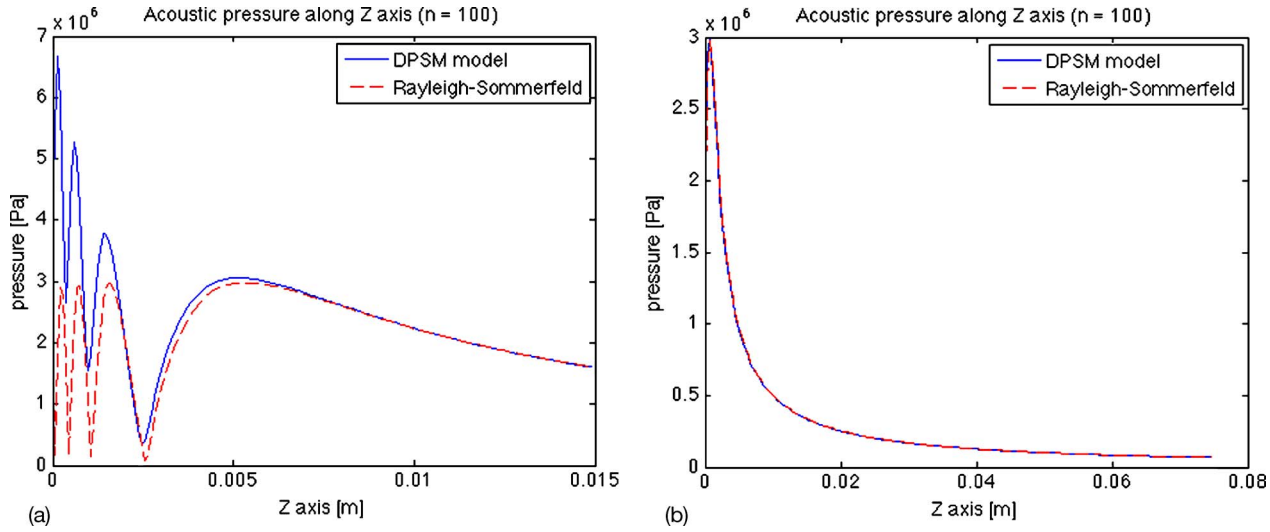


FIG. 2. (Color online) Pressure along the central axis of a 2.54 mm diameter transducer vibrating at (a) 5 MHz and (b) 1 MHz frequencies. Continuous lines are obtained from DPSM assuming uniform point source strengths [Eq. (12)] for all 100 point sources, dashed curves are RSI solutions [Eq. (13)]. Clearly, 100 point sources are not enough for modeling the near field of the 5 MHz transducer.

continuous across the transducer-fluid interface (which should be the case for viscous fluids) then from Eq. (1),

$$\mathbf{V}_S = [\mathbf{V}_{SX}, \mathbf{V}_{SY}, \mathbf{V}_{SZ}]^T = \mathbf{M}_{TSS}\mathbf{A}_S = [\mathbf{0}, \mathbf{0}, \mathbf{V}_{S0}]^T,$$

$$\therefore \mathbf{A}_S = [\mathbf{M}_{TSS}]^{-1}[\mathbf{0}, \mathbf{0}, \mathbf{V}_{S0}]^T. \quad (10)$$

Note that in Eq. (10) $\mathbf{A}_S$ has m number of unknowns, whereas $3m$ equations must be satisfied at m boundary points if all three components of velocity are to be continuous across the interface. To make the number of unknowns equal to the number of equations, three point sources are associated with every boundary point. An imaginary sphere is assumed to touch the transducer face at one point where the interface (or boundary) conditions are to be satisfied, as shown in Fig. 1. Within every sphere, three point sources that contribute to the velocity and pressure field computation at the boundary point are considered. These sources are called triplet sources.[21] With triplet source modeling one should have $3m$ equations and $3m$ unknowns. When triplet sources are considered in Eq. (10), then the dimension of both $[\mathbf{M}_{TSS}]$ and $[\mathbf{M}_{TSS}]^{-1}$ becomes $3m \times 3m$.

Instead of the velocity continuity if the pressure continuity across the interface is to be satisfied then from Eq. (1) one obtains,

$$\mathbf{A}_S = [\mathbf{Q}_{SS}]^{-1}\mathbf{P}_{S0}. \quad (11)$$

Source strengths can be computed from Eq. (9) and (10), or (11). Alternately, uniform source strengths over the transducer surface can be assumed, which requires no matrix inversion,

$$\mathbf{A}_S = \begin{bmatrix} V_{S0} & V_{S0} & V_{S0} & \cdots & V_{S0} \end{bmatrix}^T. \quad (12)$$

In Eq. (12), V_{S0} is the transducer surface velocity. Note that when the point source strengths are assumed to be equal or proportional to the transducer velocity, as shown in Eq. (12), the DPSM technique is reduced to the discretized form of the RSI representation that will be discussed in detail later.

Ultrasonic fields in front of a circular transducer face are computed by the DPSM formulation using four different Eqs. (9)–(12) and are compared with the RSI solutions in the following section. It should be noted here that the cylindrical transducer is modeled here for comparing the DPSM predicted results with analytical solutions along the central axis (z-axis in Fig. 1). Any other transducer geometry, such as transducers with elliptical, rectangular, or triangular face, can be modeled in the same manner by DPSM technique but for such geometries analytical solutions are not available.

III. NUMERICAL RESULTS

Ultrasonic pressure field in front of a circular transducer along the central axis (z-axis in Fig. 1) of the transducer is generated using Eq. (12), and the computed results are compared with the closed form expression of the central axis pressure obtained from the RSI of type II as given in Eq. (13),

$$p(z) = \rho c_f v_0[\exp(ik_f z) - \exp(ik_f\sqrt{z^2 + a^2})]. \quad (13)$$

In Eq. (13), ρ is the fluid density, c_f is the P-wave speed in the fluid, v_0 is the transducer surface velocity, z is the distance of the point from the transducer face, $k_f(=\omega/c_f)$ is the wave number in the fluid, ω is the angular frequency of the signal, and a is the radius of the transducer. Numerical computations are carried out for 2.54 and 4 mm diameter transducers. Acoustic wave speed in water is 1.49 km/s.

Figure 2 shows the pressure computed along the central axis of a 2.54 mm diameter transducer for (a) 5 MHz and (b) 1 MHz signal frequencies. Continuous lines are obtained with 100 point sources placed behind the transducer face (as shown in Fig. 1) when uniform point source strengths [Eq. (12)] are assumed. Dashed curves are obtained from Eq. (13). Agreement between the two curves is excellent for 1 MHz frequency; however, it is not so good for 5 MHz frequency in the near field (for points near the transducer, small values of z). When the number of point sources is increased

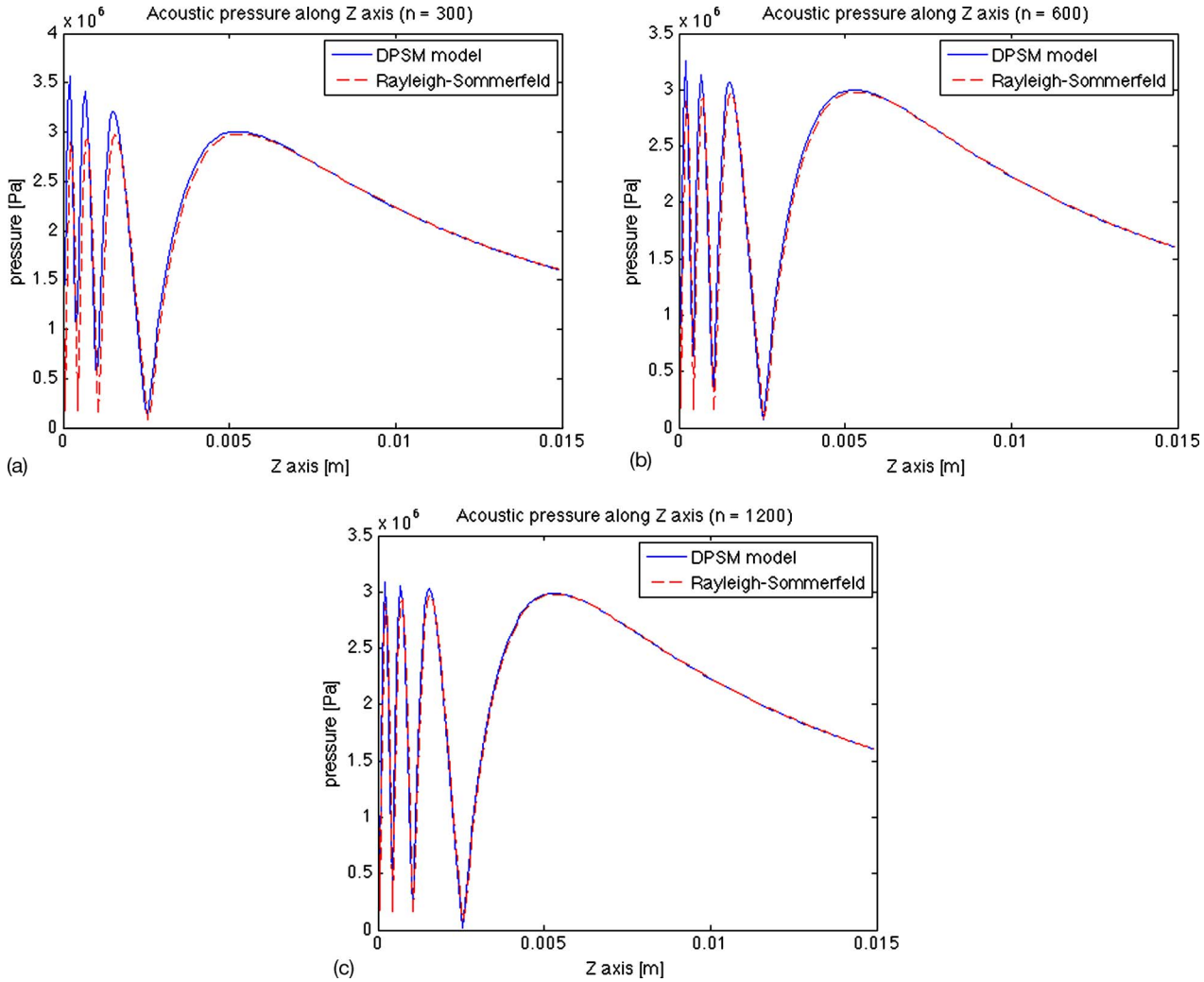


FIG. 3. (Color online) Pressure computed along the central axis of a 2.54 mm diameter transducer vibrating at 5 MHz frequency. Continuous lines are obtained from DPSM [Eq. (12)] with (a) 300, (b) 600, and (c) 1200 point sources of uniform strength. Dashed lines are RSI solutions [Eq. (13)].

from 100 to 300, 600, and 1200, then the matching between the DPSM predictions and Eq. (13) is improved significantly for the 5 MHz signal, as shown in Fig. 3.

Next, it is investigated how the DPSM results change if the point source strengths are obtained from Eq. (9). Note that Eq. (9) corresponds to the uniform V_z component on the transducer face, and the continuity of only this component of displacement is enforced across the fluid-solid interface. Figure 4 is generated with 600 point sources. Note that Figs. 3(b) and 4 are identical. This result is expected because uniform normal velocity on the transducer face implies uniform source strengths of the point sources, as is evident from the expression of the type II RSI given later.

When the transducer diameter is increased to 4 mm, then more point sources (1500) are used for keeping the λ/p ratio same for the two transducers, where λ is the wavelength and p is the pitch or distance between two neighboring point sources. Figure 5 also shows good matching between the DPSM and RSI solutions for both uniform source strength [Fig. 5(a)] and uniform normal velocity [Fig. 5(b)] cases. As expected, the near-field region is extended as the transducer diameter is increased.

Next, it is investigated how the computed field changes when all three components of velocity are assumed to be

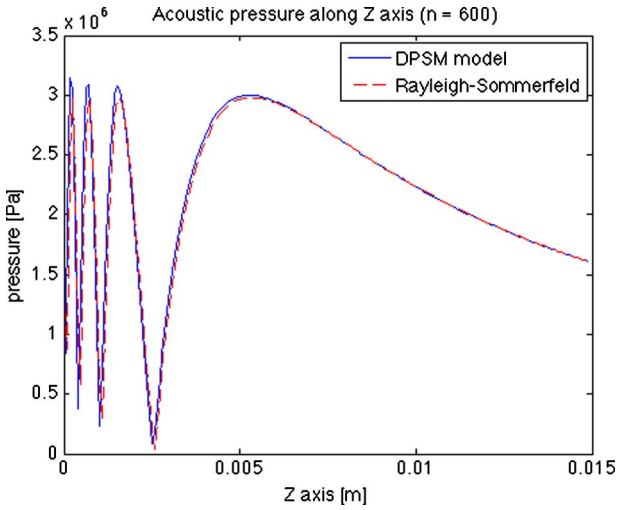


FIG. 4. (Color online) Pressure computed along the central axis of a circular transducer of 2.54 mm diameter vibrating at 5 MHz frequency. Continuous line is obtained from DPSM [Eq. (9)] with 600 simple point sources that satisfy the continuity of normal velocity across the fluid-solid interface. Dashed line is the RSI solution [Eq. (13)].

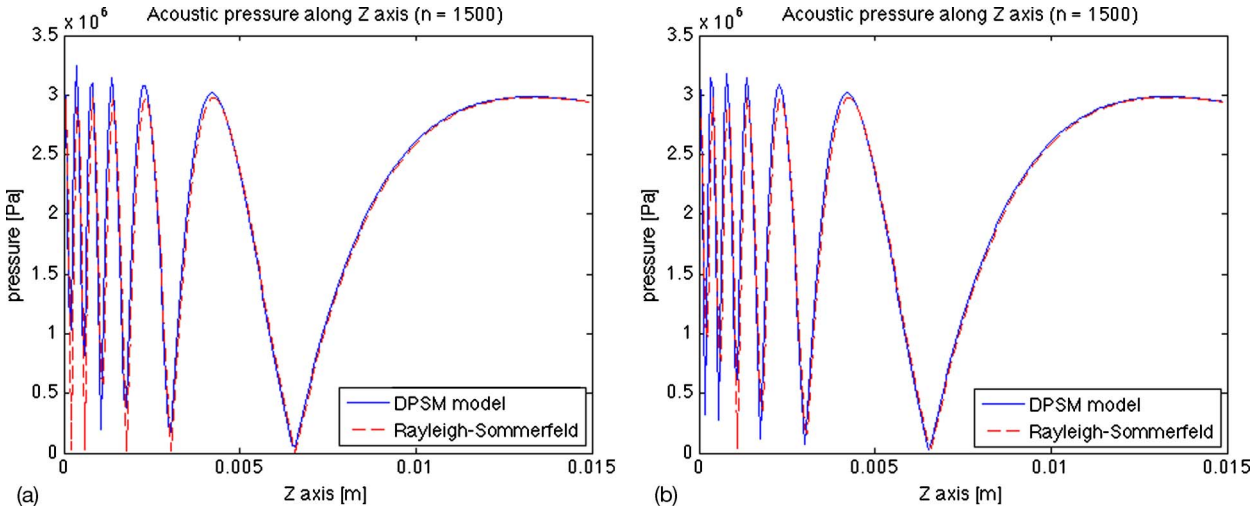


FIG. 5. (Color online) Pressure computed along the central axis of a circular transducer of 4 mm diameter vibrating at 5 MHz frequency. Dashed lines are RSI solutions obtained from Eq. (13). Continuous lines are DPSM solutions obtained from (a) Eq. (12), which assumes uniform source strengths, and (b) Eq. (9), which assumes the continuity of uniform normal velocity across the interface

continuous across the interface. In this case, the velocity of the solid and fluid particles adjacent to the transducer face should be V_{S0} in the normal direction and 0 in the other two directions. The point source strengths for this case are obtained from Eq. (10), and the results are plotted in Fig. 6 for circular transducers of 2.54 and 4 mm diameters. A clear mismatch between the DPSM and RSI solutions is evident in Fig. 6. Some mismatch is expected because the interface conditions are different for these two cases. Note that for viscous fluids no-slippage condition at the transducer face-fluid interface is expected. Therefore, one can conclude from these results that if the fluid has a small viscosity then the peak pressure should be reduced. It should be noted here that while generating Fig. 6, in the Green's function expression the fluid viscosity coefficient is not considered. Therefore, these results are only approximately true for fluids with low viscosity. Although these results are not exact solutions for viscous fluids, these plots are useful to predict the expected trend for viscous fluids. For example, from this figure one can state that in presence of a small viscosity in the fluid the pressure field is slightly reduced and the response near the transducer face is affected more by the fluid viscosity.

To investigate the effect of the uniform pressure on the computed ultrasonic field, the source strengths are obtained from Eq. (11). The pressure field on the central axis is plotted in Fig. 7 for 2.54 mm diameter transducer. It should be noted here that uniform pressure on the transducer face does not imply uniform velocity field or point source strengths on the transducer face. Type I RSI solution corresponds to the uniform pressure condition, which is the resilient disk case. Mellow[27] provided a simplified analytical solution for the central axis pressure field for this case. This solution is

$$p = p_0\left[e^{-ik_f z} - \frac{z}{\sqrt{z^2+a^2}}e^{-ik_f\sqrt{z^2+a^2}}\right], \tag{14}$$

where p_0 is the pressure on the transducer face, z; a and k_f have been defined in the text after Eq. (13). This analytical

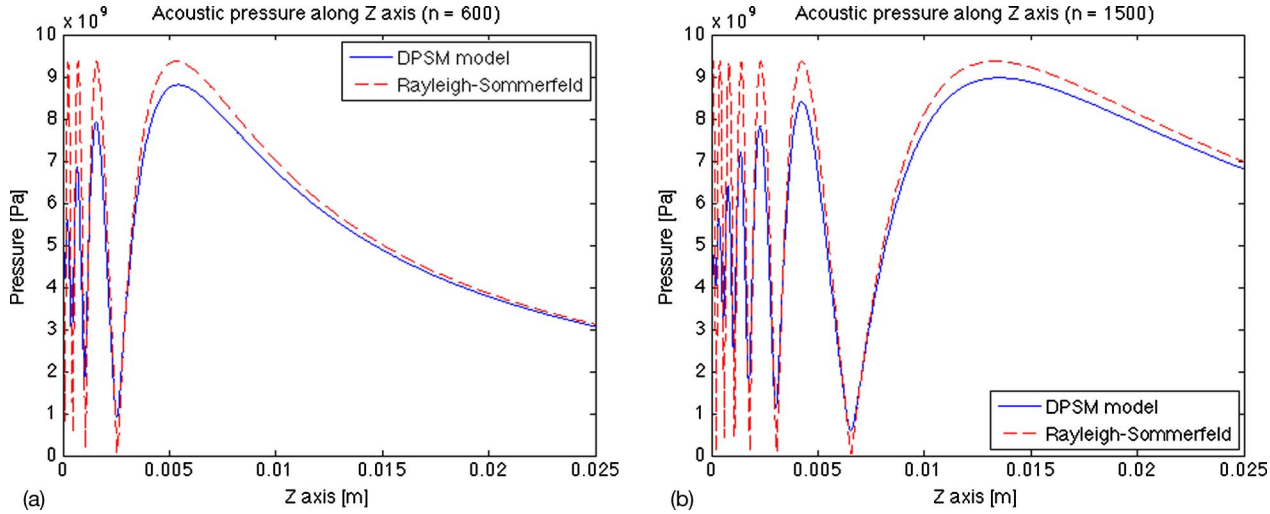


FIG. 6. (Color online) Pressure computed along the central axis for 5 MHz signal frequency. Dashed lines are RSI solutions obtained from Eq. (13). Continuous lines are DPSM solutions [Eq. (10), continuity of all three components of displacement across the interface is satisfied]. (a) 2.54 mm diameter transducer modeled by 600 triplet point sources and (b) 4 mm diameter transducer modeled by 1500 triplet point sources.

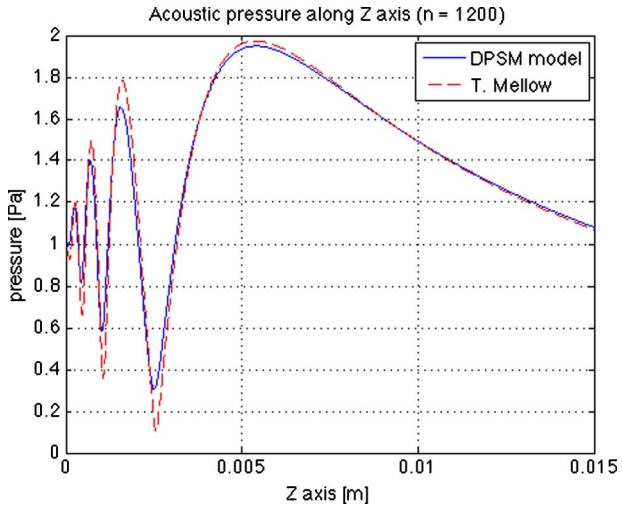


FIG. 7. (Color online) Pressure computed along the central axis of a circular transducers of 2.54 mm diameter vibrating at 5 MHz frequency. The continuous line is obtained from the DPSM analysis [Eq. (11)] assuming uniform pressure at the fluid-solid interface; dashed line is Mellow's solution [Eq. (14)].

solution is plotted in Fig. 7 as a dotted curve. DPSM solution matches with Mellow's analytical solution. However, exact matching is not expected since Mellow's solution is for a dipole pressure source in a free space, while the present solution is given for a monopole. In spite of this difference, the matching between these two results is convincing.

Note that in Fig. 7 the amplitude of the oscillating pressure along the central axis gradually increases in the near-field region as the observation point moves away from the transducer face. Similar behavior is observed when all three components of the velocity are assumed to be uniform and continuous across the interface (see Fig. 6). However, when only the normal velocity component is assumed to be uniform, then the amplitude of the oscillating pressure remains almost constant in the near-field region as the observation point moves away from the transducer face (see Fig. 5). Figures 2–5 and 7 show that RSI solutions (both types I and II) match with the DPSM results.

After computing the ultrasonic field along the central axis of the transducer for different interface continuity conditions (Figs. 2–7), it is investigated how the complete field in front of the transducer is affected by the various interface conditions. It should be noted here that although closed form analytical solutions are available along the central axis and are used for checking the DPSM results, no such closed form expression exists for the complete field. A number of recent investigations[26–29] provided solutions of Rayleigh integrals of both types (I and II) with elegant mathematical steps, but none of these solutions can produce closed form expressions for the complete ultrasonic field, and therefore dependence on semi-analytical techniques such as DPSM is unavoidable. For circular transducer geometries alternative methods[26–29] that involve summation of a finite number of terms in an infinite series are available for the ultrasonic field computation. However, for more complex transducer geometries such as a triangular, rectangular, or square transducer no analytical solution exists, and therefore, semi-analytical DPSM technique is more desirable for solving such problems. Figures 8(a)–8(d) show the total fields for four different conditions: (a) point sources have uniform strength, (b) only normal component of the velocity is assumed to be uniform (V_{S0}) on the transducer face and its continuity is satisfied across the interface, (c) all three components of the velocity are assumed to be uniform ($V_z=V_{S0}, V_x=V_y=0$) on the transducer face and the continuity of all three components of velocity is satisfied across the interface, and (d) only the acoustic pressure is assumed to be uniform (p_0) on the transducer face and its continuity is satisfied across the interface. Both contour plots and black and white gray scale images are shown side by side. Note that all four figures look similar; however, a closer observation reveals that Figs. 8(a) and 8(b) are identical (as expected) but are slightly different from Figs. 8(c) and 8(d). The conventional way of computing the ultrasonic field in front of a transducer using the RSI solution, as given in Eq. (15), does not have the same degree of flexibility in modeling different types of fluid-solid interface conditions as considered in this paper,

$$p(\mathbf{x},\omega) = \frac{-i\omega\rho}{2\pi}\int_S v_z(\mathbf{y},\omega)\frac{\exp(ikr)}{r}dS(\mathbf{y}). \tag{15}$$

Equation (15) is RSI type II, which is more popular in the nondestructive evaluation community. In this equation, $\mathbf{x}$ is the position vector where the pressure field is computed, $\mathbf{y}$ indicates the transducer surface position, and S is the transducer surface area.

IV. CONCLUSION

Ultrasonic pressure field in front of a circular transducer is obtained for different fluid-solid interface conditions. Various interface conditions such as the continuity of only one component of velocity or continuity of all three components of velocity across the fluid-solid interface, uniform velocity, or uniform pressure on the transducer face are modeled using the recently developed semi-analytical technique called DPSM. Results show some variations in the computed fields as the continuity conditions across the fluid-solid interface between the transducer and the homogeneous fluid are changed. RSI (type II) solution is limited to the case of given source strength or velocity distribution on the transducer face [Eq. (15)]. Type I RSI solution can compute the ultrasonic field in front of the transducer for a known distribution of the pressure field (or derivative of velocity) on the transducer surface.[26,27,30] When only the normal component of the velocity is assumed to be uniform and continuous across the interface, then the computed ultrasonic field is identical to the RSI solution. However, when all three velocity components are assumed to be continuous across the interface (a nonzero constant value in the normal direction and zero values parallel to the interface), then the computed field differs from the RSI. This is an inconsistent solution since no-slip condition is possible only for viscous fluids, but in the Green's function expressions the fluid viscosity has been ignored. However, these results are still presented to show the expected trend on the pressure variations in a fluid with low viscosity.

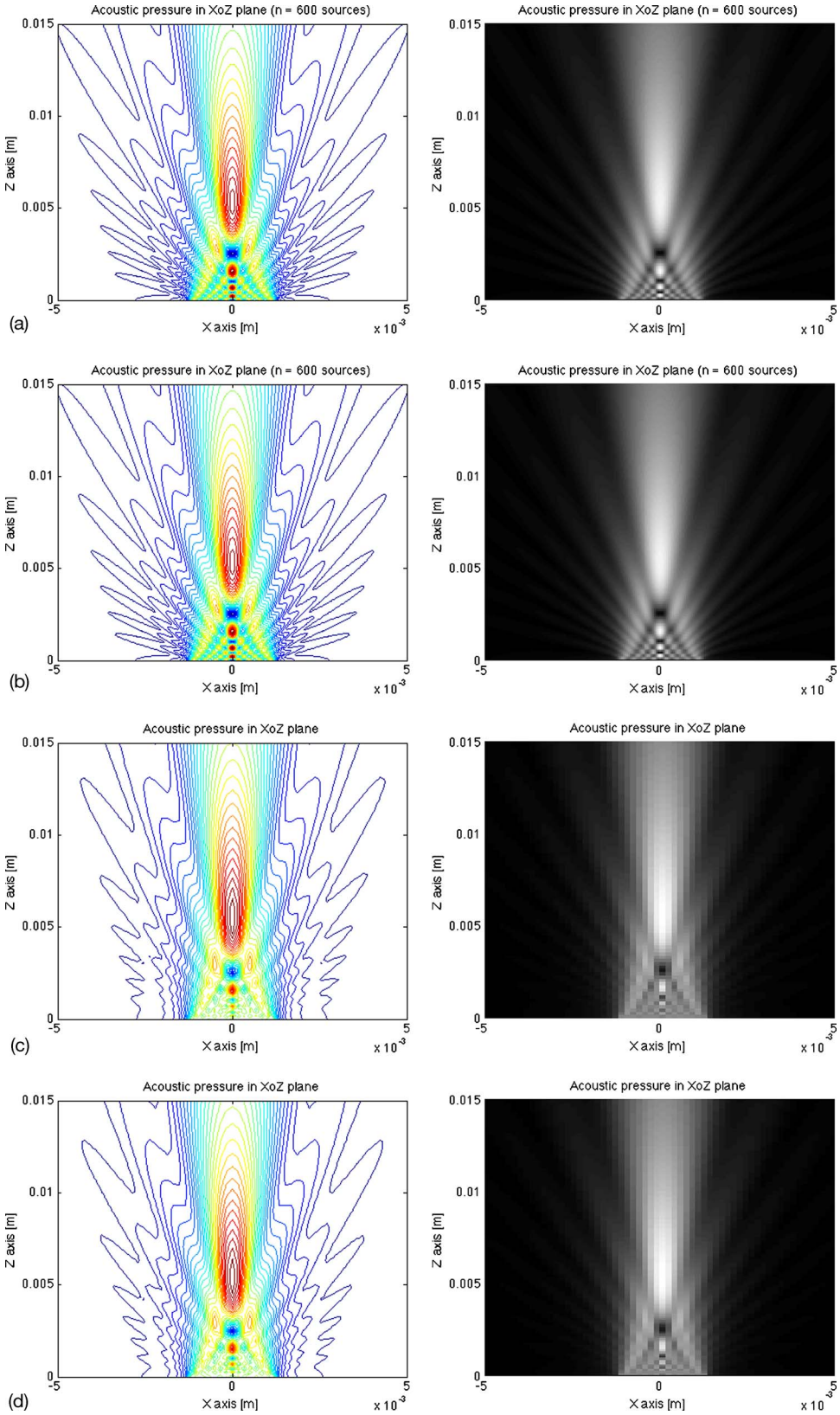


FIG. 8. (Color online) Ultrasonic pressure fields in front of a circular transducer of 2.54 mm diameter vibrating at 5 MHz frequency are modeled by 600 point sources and shown as contour plots (left figures) and in gray scale images (right figures) when (a) the point sources are assumed to have uniform strength, (b) the transducer face has uniform normal velocity component that is continuous across the interface, (c) the transducer face has uniform velocity (V_{S0} in the normal direction and 0 in the other two orthogonal directions) and all three velocity components are continuous across the interface, and (d) the transducer face produces uniform pressure at the fluid-solid interface.

The conclusion of this study can be summarized in the following manner. The widely used RSI approach gives the solution for a specific case—known source strength or known pressure field over the transducer face. RSI solution on the central axis matches well with the DPSM solution for both uniform velocity and uniform pressure cases. When all three components of velocity are assumed to be continuous across the interface, then the RSI solution differs from the DPSM solution both in the near-field and in the far-field regions. If the closed form solution of the RSI of Eq. (15) exists, then it is faster; however, if it does not exist, then the numerical integration of Eq. (15) and the DPSM technique take approximately the same computation time. This paper demonstrates the flexibility of the DPSM technique in modeling ultrasonic fields for different interface conditions.

ACKNOWLEDGMENT

This research was partially supported by the National Science Foundation Grant Nos. CMMI-0530991 and OISE-0352680.

[1]L. Rayleigh, *Theory of Sound, II* (Dover, New York, 1965), pp. 162–169.
[2]P. M. Morse and U. K. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[3]L. W. Schmerr, *Fundamental of Ultrasonic Nondestructive Evaluation—A Modeling Approach* (Plenum, New York, 1998).
[4]*Ultrasonic Nondestructive Evaluation: Engineering and Biological Material Characterization*, edited by T. Kundu (CRC, Boca Raton, FL, 2004).
[5]G. R. Harris, "Review of transient field theory for a baffled planar piston," J. Acoust. Soc. Am. **70**, 10–20 (1981).
[6]K. Sha, J. Yang, and W.-S. Gan, "A complex virtual source approach for calculating the diffraction beam field generated by a rectangular planar source," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 890–895 (2003).
[7]P. R. Stepanishen, "Transient radiation from piston in an infinite planar baffle," J. Acoust. Soc. Am. **49**, 1627–1638 (1971).
[8]J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrary shaped, apodized, and excited ultrasound transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 262–267 (1992).
[9]F. Ingenito and B. D. Cook, "Theoretical investigation of the integrated optical effort produced by sound field radiated from plane piston transducers," J. Acoust. Soc. Am. **45**, 572–577 (1969).
[10]J. C. Lockwood and J. G. Willette, "High-speed method for computing the exact solution for the pressure variations in the near field of a baffled piston," J. Acoust. Soc. Am. **53**, 735–741 (1973).
[11]G. Scarano, N. Denisenko, M. Matteucci, and M. Pappalardo, "A new approach to the derivation of the impulse response of a rectangular piston," J. Acoust. Soc. Am. **78**, 1109–1113 (1985).
[12]Z. G. Hah and K. M. Sung, "Effect of spatial sampling in the calculation of ultrasonic fields generated by piston transducers," J. Acoust. Soc. Am. **92**, 3403–3408 (1992).
[13]P. Wu, R. Kazys, and T. Stepinski, "Analysis of the numerically implemented angular spectrum approach based on the evaluation of two-dimensional acoustic fields. Part I. Errors due to the discrete Fourier transform and discretization," J. Acoust. Soc. Am. **99**, 1139–1148 (1995).
[14]T. P. Lerch, L. W. Schmerr, and A. Sedov, "Ultrasonic beam models: An edge element approach," J. Acoust. Soc. Am. **104**, 1256–1265 (1998).
[15]L. Azar, Y. Shi, and S. C. Wooh, "Beam focusing behavior of linear phased arrays," NDT & E Int. **33**, 189–198 (2000).
[16]F. Buiochi, O. Martinez, L. G. Ullate, and F. M. Espinosa, "3D computational method to study the focal laws of transducer arrays for NDE applications," Ultrasonics **42**, 871–876 (2004).
[17]S. J. Song and C. H. Kim, "Simulation of 3-D radiation beam patterns propagated through a planar interface from ultrasonic phased array transducers," Ultrasonics **40**, 519–524 (2002).
[18]R. Ahmad, T. Kundu, and D. Placko, "Modeling of phased array transducers," J. Acoust. Soc. Am. **117**, 1762–1776 (2005).
[19]J. S. Pstk, S. J. Song, and H. J. Kim, "Calculation of radiation beam field from phased array ultrasonic transducers using expanded multi-Gaussian beam model," *Advances in Safety and Structural Integrity 2005* (Sci-Tech, Vaduz, Liechtenstein, Germany, 2006), Vol. **110**, pp. 163–168.
[20]X. Zhao and T. Gang, "A study on nonparaxial beam model and ultrasonic measurement model of phased arrays," Ultrasonics (in press, 2009).
[21]D. Placko and T. Kundu, "Modeling of ultrasonic field by distributed point source method," in *Ultrasonic Nondestructive Evaluation: Engineering and Biological Material Characterization*, edited by T. Kundu (CRC, Boca Raton, FL, 2004), Chap. 2, pp. 143–202.
[22]*DPSM for Modeling Engineering Problems*, edited by D. Placko and T. Kundu (Wiley, Hoboken, NJ, 2007).
[23]S. Banerjee and T. Kundu, "Elastic wave propagation in sinusoidally corrugated waveguides," J. Acoust. Soc. Am. **119**, 2006–2017 (2006).
[24]S. Banerjee, T. Kundu, and N. A. Alnuaimi, "DPSM technique for ultrasonic field modelling near fluid-solid interface," Ultrasonics **46**, 235–250 (2007).
[25]S. Das, C. M. Dao, S. Banerjee, and T. Kundu, "DPSM modeling for studying interaction between bounded ultrasonic beams and corrugated plates," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **54**, 1860–1872 (2007).
[26]T. J. Mellow, "On the sound field of a resilient disk in an infinite baffle," J. Acoust. Soc. Am. **120**, 90–101 (2006).
[27]T. J. Mellow, "On the sound field of a resilient disk in free space," J. Acoust. Soc. Am. **123**, 1880–1891 (2008).
[28]T. D. Mast and F. Yu, "Simplified expansions for radiation from a baffled circular piston," J. Acoust. Soc. Am. **118**, 3457–3464 (2005).
[29]J. F. Kelly and R. J. McGough, "An annular superposition integral for axisymmetric radiators," J. Acoust. Soc. Am. **121**, 759–765 (2007).
[30]J. J. Stamnes, "Diffraction of three-dimensional waves," in *Waves in Focal Regions, Adam Hilger Series on Optics and Optoelectronics* (IOP, Bristol, England, 1986), Chap. 4, pp. 17–40.

Engineering modeling of traffic noise in shielded areas in cities

Erik M. Salomons,[a] Henk Polinder, Walter J. A. Lohman, Han Zhou,
Hieronymous C. Borst, and Henk M. E. Miedema
TNO Built Environment and Geosciences, Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands

(Received 14 April 2009; revised 21 August 2009; accepted 28 August 2009)

A computational study of road traffic noise in cities is presented. Based on numerical boundary-element calculations of canyon-to-canyon propagation, an efficient engineering algorithm is developed to calculate the effect of multiple reflections in street canyons. The algorithm is supported by a room-acoustical analysis of the reverberant sound fields in the source and receiver canyons. Using the algorithm, a simple model for traffic noise in cities is developed. Noise maps and exposure distributions of the city of Amsterdam are calculated with the model, and for comparison also with an engineering model that is currently used for traffic noise impact assessments in cities. Considerable differences between the two model predictions are found for shielded buildings with day-evening-night levels of 40–60 dB at the facades. Further, an analysis is presented of level differences between the most and the least exposed facades of buildings. Large level differences are found for buildings directly exposed to traffic noise from nearby roads. It is shown that by a redistribution of traffic flow around these buildings, one can achieve low sound levels at quiet sides and a corresponding reduction in the percentage of highly annoyed inhabitants from typically 23% to 18%. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3238238]

PACS number(s): 43.50.Vt, 43.28.Js, 43.50.Rq, 43.50.Qp [VEO] Pages: 2340–2349

I. INTRODUCTION

Traffic noise in cities affects the lives and health of a large number of people.[1–4] To regulate and control the effects of urban traffic noise, the European Commission requires major EU cities to produce noise maps and corresponding noise-exposure distributions of their inhabitants.[5] The first round of noise mapping took place in 2007/2008 and the next round will be in 2012. The noise-exposure distribution of a city is calculated from building exposures and numbers of people living in the buildings. Building exposure is represented by the day-evening-night level L_{den} at the *most exposed facade* of the building, indicated as $L_{den,max}$ in this article. From the levels $L_{den,max}$, the number of annoyed or highly annoyed people can be estimated by means of exposure-response relations, which are based on large statistical noise-annoyance surveys.[4,6,7] Thus, traffic noise annoyance is assessed from a single noise level per building, the level $L_{den,max}$.

Traffic noise annoyance is a complex phenomenon, however, and estimating the annoyance from the single parameter $L_{den,max}$ should be considered as an approximate, first-order approach. There are various other parameters that affect traffic noise annoyance, for example, (i) the acoustic isolation of buildings, (ii) the sound levels at the other facades, and (iii) the presence or absence of a quiet area (such as a park) near a building.[8] In this article, we focus on item (ii), in particular, on the presence or absence of a *quiet side* of a building,[8,9] which is also recognized by the European Commission as an important element of noise annoyance in cities.[5] We present an analysis of sound levels in the city of Amsterdam and investigate level differences $Q=L_{den,max}-L_{den,min}$, where $L_{den,min}$ is the level at the *least exposed facade* of a building. With increasing level difference Q at constant $L_{den,max}$, the average noise annoyance is expected to decrease.

A problem with the EU noise maps is that there is not (yet) a standard European calculation model for the sound levels. During the past decade, practical engineering models for environmental noise have been developed in the European project Harmonoise[10] and the Scandinavian project Nord2000,[11,12] which may be considered as starting points for a future European standard model. For traffic noise in cities, however, these engineering models require further development to account for multiple reflections and diffractions by buildings. Such a development should take advantage of the growing literature on model studies and experiments of urban traffic noise, reported by researchers from Sweden[13–21] and other European countries.[22–28]

In lack of a European standard model, various national calculation models have been used for the EU noise maps of 2007/2008, which makes the comparison of the noise maps rather ambiguous. One may expect that different models more or less agree on sound levels at directly exposed facades of buildings. Model differences are probably considerably larger for shielded areas such as a quiet side of a building, i.e., a side where sound waves arrive indirectly by reflection and diffraction. In this article we illustrate this by comparing noise levels and exposure distributions for Amsterdam calculated with two different models: (i) the Dutch standard model (DSM) for road traffic noise,[29] which is similar to the international ISO model,[30] and (ii) a new simple model for urban sound propagation partly based on previous work by researchers from Sweden.[13,14] It is argued that in spite of its simplicity the new model yields more realistic

[a]Author to whom correspondence should be addressed. Electronic mail: erik.salomons@tno.nl

results for quiet facades than the Dutch standard model does, by comparison with accurate reference results calculated with a numerical model based on the boundary element method[31] (BEM) for some typical urban traffic noise situations.

In general, studies of annoyance and health effects caused by traffic noise (traffic noise impact assessments) are largely based on sound levels calculated with engineering models.[1–3,5–9] The accuracy of calculated levels affects directly the results of these studies. Because of the increasing interest in the positive effects of quiet areas in cities, it is of great importance that the engineering models are "updated" to achieve accurate sound levels in quiet areas.

It has been reported before that current engineering models for traffic noise underestimate sound transfer from a street to a shielded area, i.e., canyon-to-canyon propagation, with deviations up to 10 dB.[15,27] In general, the sound level in a shielded area is determined by contributions from traffic in a large part of the city (this is explained in Sec. II), so it is not *a priori* obvious how the underestimation of canyon-to-canyon propagation affects the overall level in a shielded area, and also how it affects the overall distribution of sound levels in a city. We decided to investigate this by means of complete calculations of noise maps and exposure distributions of Amsterdam with the two models mentioned before, including about 154 000 buildings and 9000 road segments. It should be noted that the objective of the study was *not* to develop "the final model" for traffic noise in cities. Rather, the new model has been kept deliberately as simple as possible for the purpose of the study, and should only be considered as a model that is considerably more realistic for shielded areas than current engineering models are. Further development and fine-tuning are undoubtedly still necessary.

II. QUALITATIVE DESCRIPTION OF TRAFFIC NOISE IN A CITY

Traffic noise is an important element of the "urban soundscape," i.e., the total distribution of environmental sounds and noises that people in a city are exposed to.[32] The spatial distribution of traffic noise in a city is determined by the (time-dependent) distribution of vehicles on the roads, the distribution of buildings in the city, and the (time-dependent) atmospheric conditions. Reflections and diffractions by buildings play an important role in the propagation of traffic noise. For the qualitative description of urban traffic noise presented in this section, it is useful to introduce a dual representation of a city: (i) either as a flat ground surface with buildings on it or (ii) alternately as a more or less flat surface at an average rooftop level with *canyons* such as streets and courtyards (see Fig. 1).

We distinguish two types of receivers:

(1) receivers exposed directly to traffic noise and
(2) receivers shielded by buildings from traffic noise.

Figure 2 shows a schematic top view of an urban area with a receiver of type 1 (receiver R1) and a receiver of type 2 (receiver R2). The average sound level at receiver R1 is dominated by noise from cars passing by at short distance.

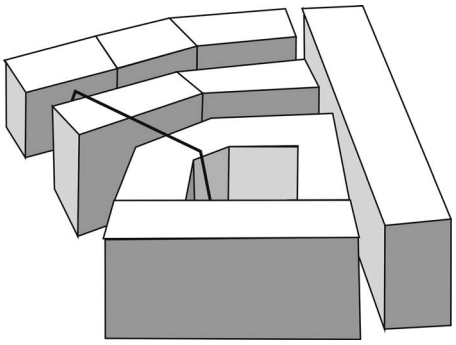


FIG. 1. Schematic representation in perspective of buildings in a city. The thick line represents a sound ray from a street canyon to a shielded canyon (courtyard).

The noise reaches the receiver by direct sound rays and by zigzag reflections in the street canyon. Contributions from cars in more distant streets, shielded by buildings, are relatively small for receiver R1. For receiver R2, however, cars in a larger region contribute to the received noise, for example, a circular region with a radius of the order of 500 m. All cars are screened by buildings and the contributions become weaker with increasing distance mainly by geometrical attenuation, i.e., spherical spreading of sound waves.

The distinction of two types of receivers more or less corresponds to the distinction of two "traffic noise fields" in a city: a direct field and a diffuse field, as introduced by Kropp *et al.*[15] The direct field is observed by a receiver that is exposed directly to noise from cars passing by, with a characteristic variation in the sound level with time. The diffuse field is most clearly observed by a receiver in a shielded area, surrounded by several roads that together determine the diffuse sound field in the shielded area. The temporal variations in the diffuse field are smaller than the temporal variations in the direct field. In practice, the sound field at a receiver is often a mixture of direct and diffuse fields.

As a preparation for a practical model for urban sound propagation presented in Sec. IV, we divide sound rays between a source and a receiver into different categories (see Fig. 2):

(c1) sound rays within a street canyon,
(c2) sound rays from a source canyon to a receiver canyon,

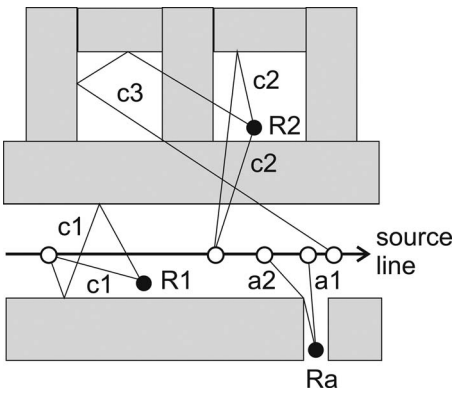


FIG. 2. Schematic top view of an urban area with a directly exposed receiver of type 1 (receiver R1) with sound rays of type c1, a shielded receiver of type 2 (receiver R2) with sound rays of types c2 and c3, and receiver Ra with sound rays a1 and a2. Open circles on the source line represent point source positions, and filled circles represent receiver positions.

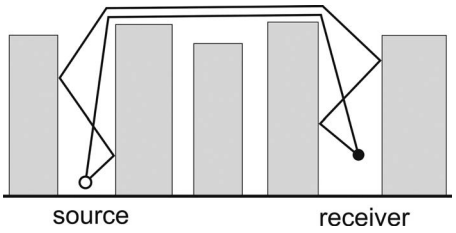


FIG. 3. Side view of two sound rays of type c2 from a source canyon to a receiver canyon.

(c3) sound rays that go from a source canyon first to an intermediate canyon and then to a receiver canyon.

Figure 3 shows a side view of two sound rays of type c2. Sound first travels upward in the source canyon to rooftop level, then it travels over several buildings and canyons, and finally it reaches the receiver canyon where sound travels downward to the receiver. Figure 3 assumes a two-dimensional geometry with infinite buildings and streets, which is a crude representation of the three-dimensional problem (see also Refs. 10 and 24). In Sec. III, we show that it is important to take into account a large number of zigzag reflections in the source and receiver canyons.

Sound rays of type c3 and higher order can in many cases be neglected for a practical model. The reason is that these rays are weaker than rays of type c2, in general, due to the additional diffractions at the intermediate canyon(s). The neglect of rays of type c3 and higher order is supported by the picture of a city as a more or less flat surface at rooftop level, with sound coming out of source canyons traveling over the elevated surface to receiver canyons (see Fig. 1). This picture is supported in Sec. III B by a room-acoustical analysis of canyon-to-canyon propagation.

It should be noted that the ray classification (c1, c2, and c3) ignores sound rays that are diffracted or reflected *around* buildings rather than *over* buildings. An example is ray a2 to receiver Ra in Fig. 2. In this example the neglect of diffracted ray a2 is not a serious problem as this ray is weaker than direct ray a1, but diffracted rays may be important in other situations without direct rays. In general, however, it has been shown that sound paths over buildings, rather than around buildings, are dominant for cities with many uninterrupted facades along streets.[14] In some cities (or city areas), however, openings between buildings along a street may play a significant role. The effect of these openings is discussed further in Sec. V.

Finally, we consider the effect of acoustic absorption of facades. As detailed information on the structure of facades is generally not available, one may adopt a practical approach and assume a reflection coefficient of 0.9 or 0.8, corresponding to a sound level reduction of 0.5 or 1 dB, respectively. This reflection attenuation represents both acoustic absorption by the facade and scattering by surface irregularities, such as windows and balconies.

III. NUMERICAL CALCULATIONS OF CANYON-TO-CANYON PROPAGATION

A practical model for urban sound propagation should be based on an efficient approach to account for the large

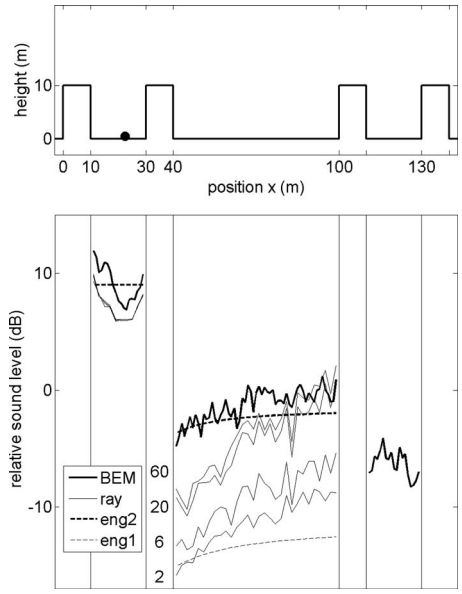


FIG. 4. Sound level relative to free field as a function of position x in the geometry shown above the graph with 10 m high buildings, calculated with a BEM model, a Fresnel-zone ray model (with 2, 6, 20, and 60 facade reflections), and engineering approaches eng1 and eng2.

number of facade reflections in street canyons.[18,27] As a preparation for such a model (presented in Sec. IV), we present in this section results of numerical calculations with a BEM model. BEM yields a numerical solution of the wave equation in the frequency domain, which implicitly includes all reflections in situations with street canyons.

A. Description and analysis of numerical calculations

The upper parts of Figs. 4 and 5 show the geometry of the BEM calculations, with street canyons between buildings with heights of 10 and 20 m, respectively. The BEM calculations have been performed for a two-dimensional system, so we assume infinite buildings and canyons. As indicated in Sec. II, the two-dimensional representation is an approximation of the real three-dimensional situation. This approxima-

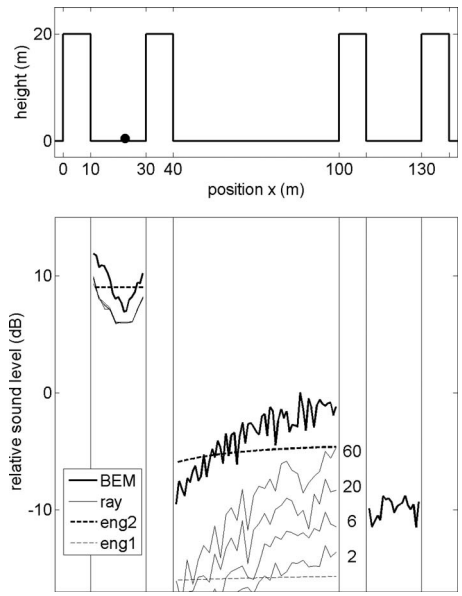


FIG. 5. As Fig. 4, for buildings with height 20 m.

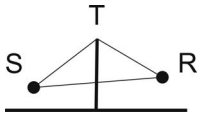


FIG. 6. Geometry for the definition of path-length difference δ_{STR}, with source S, receiver R, and barrier top T.

tion corresponds to a rotation of the buildings between the source and the receiver, such that the buildings become perpendicular to the source-receiver line (and infinitely long). For situations with barriers perpendicular to the source-receiver line, the two-dimensional representation yields an accurate approximation of the excess attenuation (i.e., the sound level relative to free field) in three dimensions.[33]

The black dot in the left canyon in Figs. 4 and 5 represents the source at height 0.5 m. To avoid large ground interference effects, we used receivers at height zero. Sound levels were integrated over a typical traffic noise spectrum in the frequency range 80–800 Hz, with relative (A-weighted) emission octave-band levels of 0, 3, and 6 dB for 125, 250, and 500 Hz, respectively. The 1 kHz contribution is not negligible for accurate absolute sound levels, but can be neglected for the present analysis: "tuning" the engineering approach for canyon-to-canyon propagation described below to BEM results. For the normalized impedance (Z) of the facades, we assumed a real value of 40, corresponding to an absorption coefficient of about 0.1. A rigid ground surface was assumed.

The lower parts of Figs. 4 and 5 show the BEM results. The graphs show the sound level relative to free field (in free field there is only geometrical attenuation and air absorption of sound waves), as a function of receiver position in the geometry indicated above the graphs. Air absorption was ignored with BEM, so calculating the sound level relative to free field required only a correction for geometrical attenuation. The sound level in the source canyon is about 9 dB above free-field level due to facade reflections and ground reflections, while in the other two canyons the relative level is about 10–15 dB lower than in the source canyon.

In the graphs we have also included results of a Fresnel-zone ray model[18,27] ("ray") and two engineering approaches ("eng1" and "eng2"). For the ray model calculations, we assumed thin walls at $x=10$, 35, and 100 m, and calculations were restricted to the region $x<100$ m. Screening attenuation by the wall at $x=35$ m was calculated with Maekawa's empirical formula[34] (which can be derived from diffraction theory[35,36]):

$$A_{\text{screen}} = 10\log(20N_F + 3), \tag{1}$$

with Fresnel number $N_F=2\delta_{STR}/\lambda$, where

$$\delta_{STR} = \pm(|ST| + |TR| - |SR|) \tag{2}$$

is the signed path-length difference illustrated in Fig. 6, and λ is the wavelength. For receivers in the shadow region behind a barrier, such as in Fig. 6, the plus sign is used. For receivers above the line of sight through S and T (the shadow boundary) the minus sign is used, up to the point where $N_F=-0.1$, so the screening attenuation decreases continuously to zero above the line of sight. In the situations of Figs. 4 and 5, the receiver is always in the shadow region, but in Sec. IV also situations with receivers above the line of sight are considered. Fresnel weighting is included in the ray model to account for the finite height of a wall as a reflector. Results are shown for sound rays with up to 2, 6, 20, and 60 facade reflections in the source and receiver canyons. Ray results deviate from the BEM results in the middle canyon even if we include rays with up to 60 facade reflections.

To describe the engineering approaches eng1 and eng2, we adopt the common engineering expression[30]

$$A_{\text{excess}} = A_{\text{ground}} + A_{\text{screen}} \tag{3}$$

for the excess attenuation (the excess attenuation is equal to *minus* the sound level relative to free field). For the ground attenuation A_{ground}, we use a constant value of −9 dB, which was fitted to the BEM result for the source canyon (for comparison, the "flat city model" described below employs a value of −6 dB; see also Sec. IV B). The screening attenuation is zero in the source canyon, while in the middle canyon we use different formulas for eng1 and eng2 to calculate the screening attenuation.

For eng1 we use Maekawa's formula (1) for a wall at $x=35$ m, with an upper limit of 25 dB for A_{screen}. The eng1 result is considerably lower than the BEM result in the middle canyon, due to the fact that zigzag reflections are not taken into account.

For eng2, we modified Maekawa's formula as follows:

$$A_{\text{screen}} = 5\log(20N_F + 3), \tag{4}$$

so we simply multiplied the screening attenuation of eng1 by a factor of 1/2, while the upper limit of 25 dB was conserved. In this way we get results closer to the BEM results in the middle canyon. Using half Maekawa's screening attenuation may be considered as a crude semi-empirical method to take into account zigzag reflections in the source and receiver canyons in an indirect way.

We performed similar calculations with BEM, ray, eng1, and eng2 for a few other situations: one with $Z=20$ instead of $Z=40$ (absorption coefficient 0.2 instead of 0.1), and one with canyon widths of 40 and 120 m instead of 20 and 60 m (see Figs. 4 and 5). The results of these calculations are not shown here, but they also indicate that approach eng2 is a reasonable engineering approach for canyon-to-canyon propagation in a city. In Sec. IV we describe an engineering model for traffic noise in a city based on approach eng2.

It is interesting to compare the BEM results and the engineering approaches eng1 and eng2 with the flat city model presented by Thorsson *et al.*[13] The flat city model employs a two-stage approach for calculating sound levels in a city. First, sound levels are calculated with all sources placed on a hard flat ground surface *without* buildings, and next the sound levels in shielded canyons are lowered by a constant "canyon-to-canyon attenuation," for which an experimental value between 6 and 10 dB is reported. This implies that the *relative* sound level in a shielded canyon is 6–10 dB lower than in a non-shielded canyon, at least for the experimental situations in Sweden on which the flat city model is based. Comparison with BEM and eng2 results, such as shown in Figs. 4 and 5, indicates that BEM and eng2

yield more realistic estimates for a city than eng1 does. A similar comparison was presented in Ref. 28. An argument for using eng2 (i.e., "half Maekawa"), rather than the constant canyon-to-canyon attenuation of the flat city model, is that one expects that higher buildings lead to lower levels in shielded canyons. It is obvious, however, that also eng2 is a crude step from eng1 to realistic levels in shielded areas. Ideally, one would like to have an expression for the canyon-to-canyon attenuation with several parameters, such as dimensions of canyons and buildings, source and receiver positions, facade structural parameters, ground parameters, and atmospheric parameters. A possible starting point for the development of such an expression is presented in Sec. III B.

B. Room-acoustical analysis

To support the engineering approach eng2, we present in this section an alternative approach to canyon-to-canyon propagation, based on the theory of room acoustics. We assume that there are so many reflections in the source and receiver canyons (see Fig. 3) that the sound fields in the canyons can be assumed as reverberant, or diffuse. The "ceilings" of the canyons can be considered as open "windows," i.e., surfaces with absorption coefficient equal to unity. Part of the sound energy generated by a source in the source canyon is absorbed by the building facades and the ground surface and part of the sound energy travels through the ceiling of the canyon. The ceiling radiates sound waves in all directions, also in the horizontal direction to the receiver canyon, where sound waves travel downward from the ceiling to the receiver. The sound power emitted by the ceiling of the source canyon is equal to WS_S/A_S, where W is the sound power of the source, S_S is the surface area of the source canyon ceiling, and $A_S=\Sigma_j S_j\alpha_j$ is the *total absorption* of the source canyon, where the "walls" (building facades, ground surface, and ceiling) have surface areas S_j and absorption coefficients α_j $(j=1,2,\ldots)$.[37] A similar energy balance holds in the receiver canyon with ceiling area S_R and absorption A_R, and one finds the following expression for the canyon-to-canyon excess attenuation:

$$A_{\text{excess}} = -10\log\left(\frac{S_S}{A_S}F\frac{S_R}{A_R}\right). \tag{5}$$

Here F is a (frequency-dependent) directivity factor accounting for the two right angles in the canyon-to-canyon sound path, one at the ceiling of the source canyon and one at the ceiling of the receiver canyon. Typically, one expects a directivity factor of the order of 0.1, corresponding to an attenuation of 10 dB.

The three factors in the argument of the logarithm in Eq. (5) correspond to the three propagation path segments: upward propagation in the source canyon, horizontal propagation over the elevated surface at rooftop level, and downward propagation in the receiver canyon. In general, absorption coefficients of building facades and ground surfaces are small compared to unity, so the factors S_S/A_S and S_R/A_R are equal to unity in good approximation. Thus, upward propagation in the source canyon and downward propagation in the receiver canyon occur with little energy loss, and the main energy loss occurs at rooftop level, so the excess attenuation is independent of building height, according to this model.

The above room-acoustical approach assumes diffuse sound fields and is therefore most appropriate for situations with high canyons and small canyon areas S_S and S_R ("chimney-like" situations). Practical situations are "somewhere between" two extreme situations: (i) the situation with completely diffuse sound fields and (ii) the situation with a single dominant sound ray without reflections and only a single diffraction at the top of the (broad) obstacle between the source and the receiver. Consequently, the "half Maekawa" approach is a reasonable approach between the "full Maekawa" approach and the opposite room-acoustical approach with excess attenuation independent of building height.

IV. MODELS

In this section, two models for traffic noise are described, the Dutch standard model[29] and a new simple model for cities based on results presented in Sec. III. Both models approximate a city as a flat ground surface with three types of elements: (i) source lines corresponding to traffic flows, (ii) buildings, and (iii) noise barriers. Buildings and barriers are represented by thin walls that reflect and diffract sound waves. A rectangular building, for example, is represented by four walls (in other words, the roof of the building is ignored).

A. Dutch standard model

The DSM for road traffic noise[29] was developed about 30 years ago, and many elements of the model have been used later for the international ISO model.[30] For the study presented here, the reader does not have to understand all elements of DSM. The elements that are relevant are described below. For the convenience of those readers who do want to understand all elements, however, an English translation of the model description has been prepared, which is available through the Electronic Physics Auxiliary Publication Service (EPAPS) of the American Institute of Physics.[29]

The model distinguishes three types of vehicles: light vehicles (passenger cars), medium-heavy vehicles (light trucks and buses), and heavy vehicles (heavy trucks). A vehicle is represented as a point source at height 0.75 m, and different (speed-dependent) octave-band emission spectra are used for the three vehicle types, including spectral correction terms for road surfaces such as porous asphalt. Numbers of vehicles per unit length are determined by vehicle intensities (numbers of vehicles per hour) and driving speeds for the three vehicle types. Day-evening-night levels L_{den} are determined from levels calculated for the day period (7–19 h), evening period (19–23 h), and night period (23–7 h),[5] using appropriate vehicle intensities for the three periods.

Source lines are divided into small segments and source points are placed at the centers of the segments. The sound level at a receiver is calculated by logarithmic summation of contributions from sound rays between source points and the receiver. The model takes into account sound rays with zero

or one facade reflection. Multiple facade reflections are (usually) ignored with DSM. Sound rays with zero facade reflections are of type c1 or c2 (see Sec. II). Sound rays with one facade reflection are of type c1, c2, or c3. Ground reflections are taken into account indirectly by a ground attenuation term, which is similar to the ground attenuation term employed by the ISO model.[30] This term is valid for downward-refracting propagation conditions, and a meteorological correction term is included to account for upward-refracting conditions, so the result is an estimate of a long-term average sound level.

If the ground projection of a sound ray intersects one or more walls, then the model takes into account a screening attenuation based on the wall with the largest path-length difference δ_{STR}, while all other walls are ignored for the screening attenuation. The screening attenuation is calculated with a formula similar to Maekawa's formula (1), with an upper limit of 25 dB (so similar to the eng1 approach of Sec. III).

It should be noted that atmospheric wind may cause considerable deviations from Maekawa's formula,[38,39] in particular, in open areas near highways. In an urban environment, however, wind effects on barrier attenuation are smaller in general (downward-refracting wind-speed gradients are larger near a single barrier in an open area than near a barrier that is close to several other barriers), and it is considered more important to take into account multiple reflections in street canyons, as described in Sec. IV B.

B. Semi-empirical model for traffic noise in cities

In this section, we present a new simple model for urban traffic noise, based on the eng2 approach of Sec. III for canyon-to-canyon propagation. The model will be referred to as street-canyon model (SCM). As the eng2 approach was constructed to achieve better agreement with BEM results in shielded areas (i.e., better than with eng1), the model can be considered as a semi-empirical model for traffic noise in cities.

Reference 28 provides some experimental support for BEM results in shielded areas, by comparison with outdoor measurements and scale-model measurements. Reference 13 presents experimental data as a basis for the flat city model, and thereby provides experimental support for the eng2 approach (see Sec. III A). Additional measurements and BEM calculations should be performed in the future to develop a more accurate or refined engineering model for canyon-to-canyon propagation. The objective of the present study was to use the approximate eng2 approach to gain more insight in sound levels in shielded areas in cities (see Sec. V).

The SCM model is identical to the DSM model except for the set of sound rays taken into account and for the calculation of the excess attenuation. The SCM model takes into account all sound rays of types c1 and c2 (in an indirect way), while rays of types c3 and higher order are neglected. All zigzag reflections in the source and receiver canyons are taken into account indirectly by using approach eng2, so no geometrical facade reflection calculations are necessary, which makes the model very efficient. For each source (point)-receiver combination, the excess attenuation is $A_{\text{ground}}+A_{\text{screen}}$, with $A_{\text{ground}}=-3$ dB and A_{screen} given by Eq. (4) based on the intersected wall with the largest path-length difference δ_{STR} (if no wall is intersected by the ground projection of the source-receiver line then the screening attenuation is zero). We use an upper limit of 20 dB for the screening attenuation, instead of the rather high value of 25 dB of DSM.

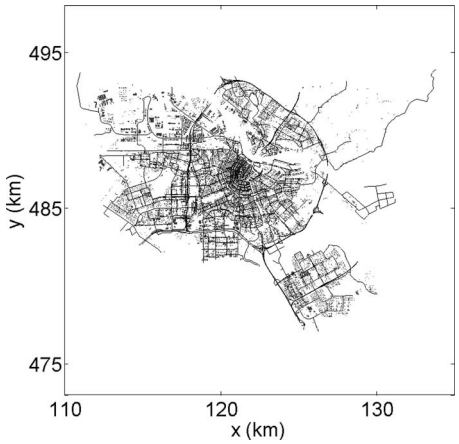


FIG. 7. Map of Amsterdam used for the calculations, with roads shown as lines and buildings as gray polygons.

The value of −3 dB for A_{ground} was simply adjusted to get agreement with DSM results for receivers directly exposed to traffic noise (see Sec. V). The value of −3 dB differs from the value of −9 dB used in Sec. III, mainly because (i) the meteorological correction term was not included in Sec. III and (ii) we focus on a receiver height of 4 m (see Sec. V) rather than receiver height zero in Sec. III. So if one wants to achieve sound levels for height 4 m and including the meteorological correction term, then a value of −3 dB should be used for A_{ground}.

V. CALCULATIONS FOR AMSTERDAM

We have performed calculations of traffic noise levels for the city of Amsterdam with the two models described in Sec. IV. The input data for the calculations comprised the following

(i) coordinates of vertices of about 154 000 buildings, 450 noise barrier segments, and 9000 road segments (minor roads with low vehicle intensities were not included),
(ii) vehicle intensities and driving speeds for the road segments (for the year 2006, based on counted numbers of vehicles and a traffic-flow model calculation),
(iii) numbers of inhabitants per building (the total number of inhabitants is about 800 000).

Figure 7 shows a map of the city, with the roads shown as lines and the buildings as gray polygons.

Figure 8 shows two sound rays included in the DSM calculation, a direct ray of type c2 and a ray with one facade reflection of type c3. Figure 9 shows a side view of the walls (facades) that are intersected by the first ray. The source is at height 0.75 m above the road surface (which is at height 1 m

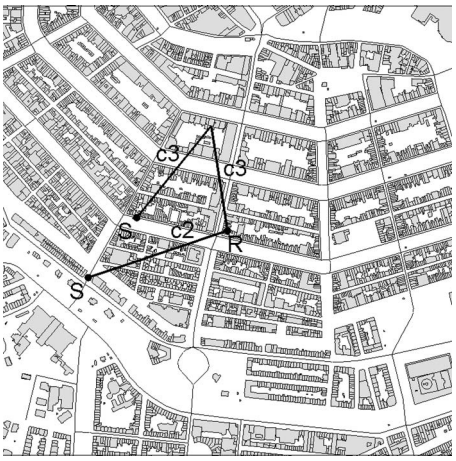


FIG. 8. Area of Amsterdam of 1×1 km^2, with two sound rays included in the DSM calculation (S=source, R=receiver), a direct ray of type c2 and a ray with one facade reflection of type c3.

here), and the receiver is at height 4 m (following Ref. 5). Heights of walls range from about 10 to 20 m here. The average building height for Amsterdam is 15 m, with a few buildings up to 80 m.

Receivers were placed at the centers of all (free) building facades, at a normal distance of 1 cm and at a height of 4 m. Reflections by the facade behind the receiver were not included in the DSM calculation, so the sound levels represent incident sound, as required for application of exposure-response relations[1,5,6] (described below). We included sources up to 600 m from the receiver, and we verified by test calculations that this truncation had a negligible effect in most cases.

Figure 10 shows spectra of the day-evening-night level for a receiver in front of a building (directly exposed receiver) and a receiver at the back of the same building (shielded receiver), calculated with SCM and DSM. Broadband day-evening-night levels are indicated as legends in the figure. For the directly exposed receiver, the DSM and SCM results are both 71 dB (in fact, the value of A_{ground} was adjusted to achieve this agreement; see Sec. IV). For the receiver at the back of the building, however, the SCM result is 7 dB higher than the DSM result.

Figure 11 shows an area of 200×200 m^2, with gray levels of buildings corresponding to the day-evening-night level at the most exposed facade ($L_{\text{den,max}}$), calculated with

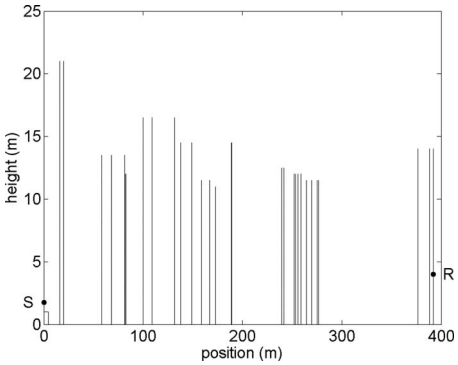


FIG. 9. Side view of the building walls (facades) for the ray of type c2 shown in Fig. 8.

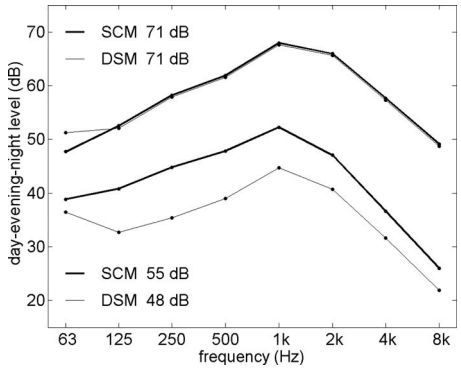


FIG. 10. Spectra of the day-evening-night level for a receiver in front of a building (directly exposed receiver) and a receiver at the back of the same building (shielded receiver), calculated with SCM and DSM. Broadband day-evening-night levels are also indicated in the figure.

the SCM model. Thin lines represent source lines. Buildings along the streets have $L_{\text{den,max}}$ levels around 70 dB, while buildings shielded by other buildings have $L_{\text{den,max}}$ levels around 55 dB.

Figure 12 shows the same area as in Fig. 11, but now the gray level represents the level difference $Q=L_{\text{den,max}}-L_{\text{den,min}}$, where $L_{\text{den,min}}$ is the day-evening-night level at the least exposed facade. The figure shows that Q is high (15–20 dB) for the buildings along the streets, while Q is low for the shielded buildings.

Figure 13 shows exposure distributions of $L_{\text{den,max}}$, i.e., the percentage of inhabitants exposed per decibel interval, calculated with SCM and DSM. The two distributions agree quite well above about 60 dB, but below 60 dB the distributions deviate from each other, due to the underestimation of sound levels in shielded areas by DSM. The distributions show two distinct maxima, one around 68 dB and one around 50 dB. The maximum around 68 dB represents directly exposed buildings, and the maximum around 50 dB represents shielded buildings.

It should be noted that the fact that minor roads with low vehicle intensities have not been included in the calculation probably has a significant effect on the lower part of the exposure distribution, say, below 45 or 50 dB. This indicates

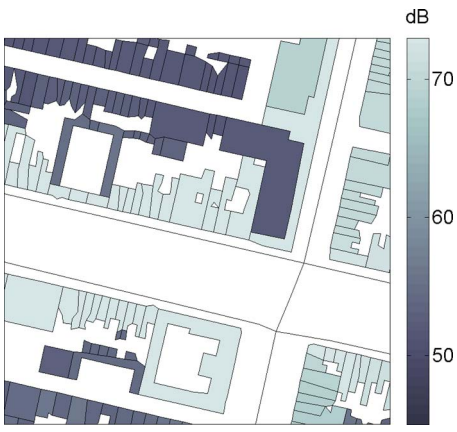


FIG. 11. (Color online) Area of 200×200 m^2, with gray levels of buildings corresponding to the day-evening-night level at the most exposed facade ($L_{\text{den,max}}$), calculated with SCM.

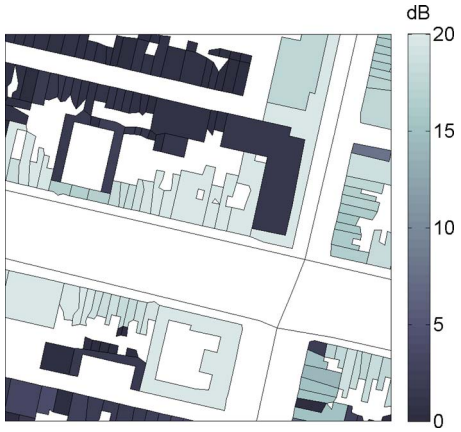


FIG. 12. (Color online) Same area as in Fig. 11, with the gray levels of buildings representing the level difference $Q=L_{\mathrm{den,max}}-L_{\mathrm{den,min}}$.

that, in addition to using an accurate sound model, it is important to have accurate and complete input data for a noise map calculation.

From the exposure distribution shown in Fig. 13, the distribution of highly annoyed inhabitants has been determined. For a given sound level $L_{\mathrm{den,max}}$ of incident sound at the most exposed facade, the percentage P of highly annoyed inhabitants is given by the relation[6]

$$P = 9.868 \times 10^{-4}(L_{\mathrm{den,max}} - 42)^3 - 1.436 \times 10^{-2}(L_{\mathrm{den,max}} - 42)^2 + 0.5118(L_{\mathrm{den,max}} - 42). \quad (6)$$

With increasing sound level $L_{\mathrm{den,max}}$, the percentage of highly annoyed inhabitants increases. At 60 dB the percentage is about 10% and at 70 dB the percentage is about 25%. Relation (6) is valid for $L_{\mathrm{den,max}} \geq 42$ dB; for $L_{\mathrm{den,max}} < 42$ dB it is assumed that P is zero. The distribution of highly annoyed inhabitants, as calculated from the exposure distribution in Fig. 13 using Eq. (6), is shown in Fig. 14. As indicated in the legend of the graph, the total percentage of highly annoyed inhabitants (i.e., the sum of the distribution over all dB intervals) is 9.6% with SCM and 8.1% with DSM.

The distributions labeled SCM(q) in Figs. 13 and 14 were calculated with SCM with a *quiet-side correction* introduced in Ref. 8 to account for the reduced annoyance when inhabitants have access to a quiet side. The quiet-side correction (in decibels) is calculated from the level difference

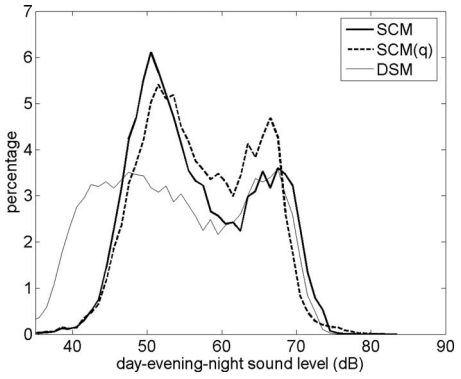


FIG. 13. Exposure distribution of $L_{\mathrm{den,max}}$, i.e., the percentage of inhabitants exposed per dB interval, calculated with SCM, DSM, and SCM(q) (SCM with quiet-side correction).

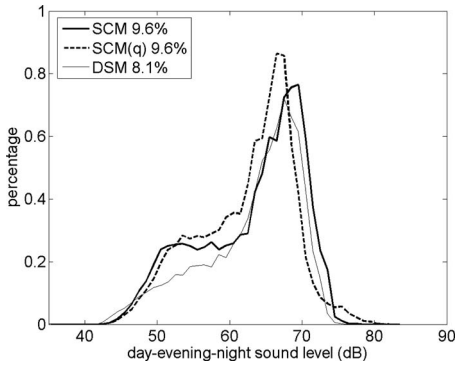


FIG. 14. Distribution of highly annoyed inhabitants corresponding to Fig. 13.

$Q=L_{\mathrm{den,max}}-L_{\mathrm{den,min}}$ introduced before, and is added to the level $L_{\mathrm{den,max}}$ at the most exposed façade. The resulting corrected level is denoted as $L_{\mathrm{den,q}}$:

$$L_{\mathrm{den,q}} = L_{\mathrm{den,max}} + [a(Q - Q_{\mathrm{av}})L_{\mathrm{den,max}} + b(Q - Q_{\mathrm{av}})]. \quad (7)$$

The term in square brackets is the quiet-side correction, with $a=-0.016$ and $b=0.7$. The values of a and b are based on approximate indications of annoyance reduction due to a quiet side.[8] The values should be improved by more extensive annoyance surveys including effects of a quiet side. The quiet-side correction is positive for low values of Q and negative for high values of Q. Consequently, a high value of Q corresponds to a corrected level $L_{\mathrm{den,q}}$ that is lower than $L_{\mathrm{den,max}}$, and consequently to a reduced percentage of highly annoyed inhabitants, as calculated with Eq. (6) from the corrected level $L_{\mathrm{den,q}}$. The quiet-side correction is zero if Q is equal to an average value Q_{av}, which is determined by the condition that the total number of highly annoyed inhabitants is equal with and without quiet-side correction (see Fig. 14), which yields $Q_{\mathrm{av}}=12.43$ dB in this case [this approach was followed since the annoyance surveys underlying Eq. (6) implicitly included the quiet-side correction]. The quiet-side correction is negative for $Q>Q_{\mathrm{av}}$ and positive for $Q<Q_{\mathrm{av}}$.

Figure 15 shows the distribution of the level difference Q, calculated with SCM. The distribution shows that Q varies between 0 and 25 dB, with a distinct maximum around 17 dB.

Figure 16 shows a graph of the sound level $L_{\mathrm{den,max}}$ at the most exposed facade as a function of the difference Q (each dot in the graph represents a building). The graph shows that high values of Q occur mainly for high levels

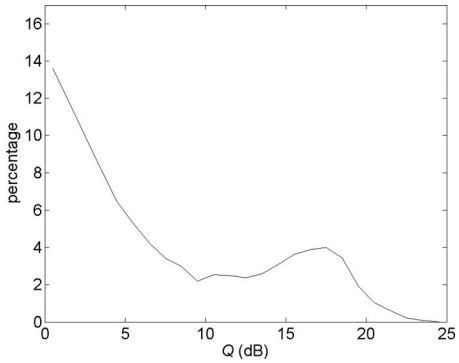


FIG. 15. Distribution of difference Q, calculated with SCM.

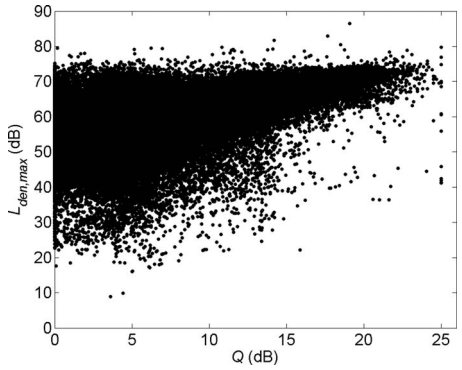


FIG. 16. Sound level $L_{den,max}$ at the most exposed facade as a function of the difference Q.

$L_{den,max}$, i.e., for buildings exposed directly to traffic noise (cf. Fig. 12). These results suggest that the maximum in the Q distribution in Fig. 15 around $Q=17$ dB corresponds to the maximum in the $L_{den,max}$ distribution in Fig. 13 around $L_{den,max}=68$ dB.

For $Q=17$ dB and $L_{den,max}=68$ dB, the quiet-side correction in Eq. (7) is about −2 dB. This corresponds to a reduction in the percentage of highly annoyed inhabitants from 21% to 18%. On the other hand, the quiet-side correction is positive for $Q=0$, i.e., for buildings that are exposed equally at all facades (at least for $L_{den,max}>43.75$ dB, which is the range of interest in Figs. 13 and 14; for $L_{den,max}<43.75$ dB one may assume that the quiet-side correction is zero). An illustration of the effect of a quiet side is given by the example shown in Fig. 17. In situation A, we have two roads with equal traffic intensities, one road at the front of the building and one road at the back. The L_{den} levels are 65 dB at both facades, so we have $L_{den,max}=65$ dB and $Q=0$. In situation B, we have removed one of the roads and doubled the traffic intensity on the other road. In other words, we have relocated all traffic to one side of the building. We have a level of 68 dB at the most exposed facade and assume a level of 51 dB at the quiet side, so we have $L_{den,max}=68$ dB and $Q=17$ dB in situation B. From Eqs. (6) and (7), we find $L_{den,q}=69.2$ dB and $P=23\%$ for situation A and $L_{den,q}=66.2$ dB and $P=18\%$ for situation B. The statistical probability that the inhabitants of the building are highly annoyed by the traffic noise is reduced from 23% to 18% by relocating the traffic to one side of the building.

The positive effect of a quiet side is significant and should be taken into account for an accurate assessment of traffic noise annoyance in a city. City planners may employ the positive effects of quiet sides and perform acoustical optimizations of the orientations of buildings with respect to roads, taking into account the quiet-side correction of the sound level at the most exposed facade. It should be recalled that the analysis presented in this article made use of the observation that sound propagation *over* buildings dominates in cities with many uninterrupted facades along streets,[14] so propagation *around* buildings through diffraction and multiple reflection was neglected. Consequently, city planners should avoid large openings between buildings along streets. The effect of openings is twofold. On the one hand, openings lead to an increase in sound levels at some receivers by

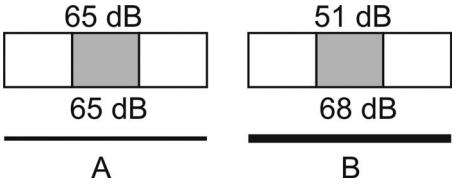


FIG. 17. Schematic representation of two situations (A and B) illustrating the effect on noise annoyance of a redistribution of traffic such that a quiet side of a building is created. Situation A: two roads (thick lines) at the front and the back of a building (gray area), with equal traffic intensities I_A (number of vehicles per hour). Situation B: single road with double traffic intensity $2I_A$. The L_{den} levels at the facades are indicated.

propagation around buildings (through the openings). On the other hand, the number of multiple reflections in a street canyon is reduced by openings, which leads to a decrease in sound levels. Further studies are required to determine how exposure distributions are modified by these two opposite effects of openings between buildings. It should be noted that decisions for avoiding openings between buildings should be based not only on noise but also on other human response factors, such as visual attractiveness of a city.

VI. CONCLUDING REMARKS

The results of this study are of interest for the interpretation of noise maps and exposure distributions of major EU cities, which were reported in 2007/2008 in the framework of the Environmental Noise Directive.[5] Part of the differences between reported noise maps may represent real differences of traffic noise exposure, but another part is probably due to the fact that different (national) noise models were used for different cities or countries. In any case, the present study has shown that the Dutch standard model,[29] which is similar to the ISO model,[30] yields unrealistically low levels in shielded areas, a result that agrees with previous published results.

Currently efforts are being undertaken for developing a standard European noise model for future EU noise mapping rounds. This development will probably be based on recent work performed in the Harmonoise and Nord2000 projects,[10–12] and may also take advantage of the results presented in this article. In particular, for noise levels in shielded areas, care should be taken to avoid unrealistically low levels.

The present study has also demonstrated that buildings that are directly exposed to traffic noise, i.e., buildings along streets, often have a *quiet side* with a sound level that is 15–20 dB lower than the level at the most exposed facade. A quiet side allows inhabitants a possibility to "escape" from the traffic noise at the most exposed facade, and thereby reduces the annoyance and possibly also sleep disturbance and negative health effects. We estimated in this article that the statistical probability that inhabitants of directly exposed buildings are highly annoyed by traffic noise can be reduced from 23% to 18% by a redistribution of traffic flow near the building such that a quiet side is created. City planners should take advantage of the positive effects of quiet sides of buildings. To shield courtyards from direct exposure to traffic noise, large openings between buildings along streets should be avoided.

The new model for traffic noise in cities presented in this article (SCM) should be considered as a first-order improvement of current engineering models DSM or ISO for shielded areas. Further improvement and fine-tuning of the model is required. Further improvement of the method for assessing the positive effect of a quiet side is also required. We hope that this work inspires readers, in particular, readers from outside Europe, to contribute to further developments of practical models for traffic noise in cities, although the topographical structure of non-European cities may differ considerably from the topographical structure of European cities.

ACKNOWLEDGMENTS

Interesting discussions with A. Randrianoelina and E. Gerretsen are gratefully acknowledged.

[1]E. Ohrström, L. Barregård, E. Andersson, A. Skånberg, H. Svensson, and P. Angerheim, "Annoyance due to single and combined sound exposure from railway and road traffic," J. Acoust. Soc. Am. **122**, 2642–2652 (2007).
[2]H. A. Nijland, S. Hartemink, I. van Kamp, and B. van Wee, "The influence of sensitivity for road traffic noise on residential location: Does it trigger a process of spatial selection?," J. Acoust. Soc. Am. **122**, 1595–1601 (2007).
[3]Y. de Kluizenaar, R. T. Gansevoort, H. M. E. Miedema, and P. E. de Jong, "Hypertension and road traffic noise exposure," J. Occup. Environ. Med. **49**, 484–492 (2007).
[4]D. Ouis, "Annoyance caused by exposure to road traffic noise: An update," Noise Health **4**, 69–79 (2002); URL: http://www.noiseandhealth.org (Last viewed March 28, 2009).
[5]European Directive on Environmental Noise, 2002/49/EC, URL: http://ec.europa.eu/environment/noise/home.htm (Last viewed March 28, 2009).
[6]H. M. E. Miedema and C. G. M. Oudshoorn, "Annoyance from transportation noise: Relationships with exposure metrics DNL and DENL and their confidence intervals," Environ. Health Perspect. **109**, 409–416 (2001).
[7]H. C. Borst and H. M. E. Miedema, "Comparison of noise impact indicators, calculated on the basis of noise maps of *DENL*," Acta Acust. Acust. **91**, 378–385 (2005).
[8]H. M. E. Miedema and H. C. Borst, "Rating environmental noise on the basis of noise maps," Report for EC project Quiet City Transport (QCity), 2007. A short version of the work presented in this report was presented at the Euronoise Conference in 2006, in Tampere, Finland (available on the CD-ROM).
[9]E. Öhrström, A. Skånberg, H. Svensson, and A. Gidlöf-Gunnarsson, "Effects of road traffic noise and the benefit of access to quietness," J. Sound Vib. **295**, 40–59 (2006).
[10]H. G. Jonasson, M. Dittrich, D. van Maercke, J. Defrance, E. Salomons, I. Noordhoek, D. Heimann, B. Plovsing, G. Watts, X. Zhang, E. Premat, I. Schmich, F. Aballea, M. Baulac, F. de Roo, M. Bakermans, D. Kühner, B. De Coensel, D. Botteldooren, F. Vanhove, S. Logghe, and R. Bütikofer, "Building Europe's future harmonized noise mapping methods," Acta Acust. Acust. **93**, 173–262 (2007); (special issue about the European projects Harmonoise and Imagine). See also URL: http://www.imagine-project.org (Last viewed March 28, 2009).
[11]J. Kragh, B. Plovsing, S. Å. Storeheier, G. Taraldsen, and H. G. Jonasson, "Nordic environmental noise prediction method. Nord2000 summary report. General Nordic sound propagation model and applications in source-related prediction methods," DELTA Acoustics & Vibration Report No. 1719/01, 2002, Lyngby, Denmark; available from URL: http://www.delta.dk/nord2000 (Last viewed March 28, 2009).
[12]G. B. Jónsson and F. Jacobsen, "A comparison of two engineering models for outdoor sound propagation: Harmonoise and Nord2000," Acta Acust. Acust. **94**, 282–289 (2008).
[13]P. Thorsson, M. Ögren, and W. Kropp, "Noise levels on the shielded side in cities using a flat city model," Appl. Acoust. **65**, 313–323 (2004).
[14]P. J. Thorsson and M. Ögren, "Macroscopic modeling of urban traffic noise—Influence of absorption and vehicle flow distribution," Appl. Acoust. **66**, 195–209 (2005).
[15]W. Kropp, J. Forssén, M. Ögren, and P. Thorsson, "The failure of traditional traffic noise control for quiet areas," in Proceedings of Internoise 2004, Prague.
[16]M. Ögren and W. Kropp, "Road traffic noise propagation between two dimensional city canyons using an equivalent sources approach," Acta Acust. Acust. **90**, 293–300 (2004).
[17]M. Ögren, "Prediction of traffic noise shielding by city canyons," Ph.D. thesis, Chalmers University, Göteborg, Sweden (2004).
[18]J. Forssén and M. Hornikx, "Statistics of A-weighted road traffic noise levels in shielded urban areas," Acta Acust. Acust. **92**, 998–1008 (2006).
[19]M. Hornikx and J. Forssén, "The 2.5-dimensional equivalent sources method for directly exposed and shielded urban canyons," J. Acoust. Soc. Am. **122**, 2532–2541 (2007).
[20]M. Hornikx and J. Forssén, "A scale model study of parallel urban canyons," Acta Acust. Acust. **94**, 265–281 (2008).
[21]M. Hornikx and J. Forssén, "Noise abatement schemes for shielded canyons," Appl. Acoust. **70**, 267–283 (2009).
[22]T. van Renterghem, E. Salomons, and D. Botteldooren, "Parameter study of sound propagation between city canyons with a coupled FDTD-PE model," Appl. Acoust. **67**, 487–510 (2006).
[23]T. van Renterghem, E. Salomons, and D. Botteldooren, "Efficient FDTD-PE model for sound propagation in situations with complex obstacles and wind profiles," Acta Acust. Acust. **91**, 671–679 (2005).
[24]T. van Renterghem and D. Botteldooren, "Numerical evaluation of sound propagating over green roofs," J. Sound Vib. **317**, 781–799 (2008).
[25]J. Kang, "Numerical modelling of the sound fields in urban streets with diffusely reflecting boundaries," J. Sound Vib. **258**, 793–813 (2002).
[26]J. Picaut, T. Le Pollès, P. L'Hermite, and V. Gary, "Experimental study of sound propagation in a street," Appl. Acoust. **66**, 149–173 (2005).
[27]E. Salomons, "A cellular automaton for urban traffic noise," in Proceedings of the Joint ASA-EAA Meeting (Acoustics 08), Paris (2008).
[28]A. Randrianoelina and E. Salomons, "Traffic noise in shielded urban areas: Comparison of experimental data with model results," in Proceedings of the Joint ASA-EAA Meeting (Acoustics 08), Paris (2008).
[29]See EPAPS Document No. E-JASMAN-126-051911 for Dutch standard model for road traffic noise ("Calculation and measurement method for road traffic noise 2002"), regulation LMV 2002 025825 of Dutch Ministry of Environment (in Dutch). The model provides a non-spectral method SRM1 and a spectral method SRM2; for this study the method SRM2 has been used, referred to as DSM (Dutch standard model) in this article. For the calculation of noise maps in the framework of the END (Ref. 5), a simplified Dutch method SKM2 has been developed; the description of SKM2 is not yet sufficiently unambiguous for use in a scientific study. The model descriptions can be downloaded from URL http://www.stillerverkeer.nl (Last viewed March 28, 2009). An English translation of the model description has been prepared, which is available through the Electronic Physics Auxiliary Publication Service (EPAPS) of the American Institute of Physics. For more information on EPAPS, see http://www.aip.org/pubservs/epaps.html.
[30]ISO 9613-2, Acoustics—Attenuation of sound during propagation outdoors—Part 2: General method of calculation, ISO, 1996. URL: http://www.iso.org (Last viewed March 28, 2009).
[31]*Boundary Element Methods in Acoustics*, edited by R. D. Ciskowski and C. A. Brebbia (Elsevier, London, 1991).
[32]D. Botteldooren, B. De Coensel, and T. De Muer, "The temporal structure of urban soundscapes," J. Sound Vib. **292**, 105–123 (2006).
[33]E. M. Salomons, R. Blumrich, and D. Heimann, "Eulerian time-domain model for sound propagation over a finite-impedance ground surface. Comparison with frequency-domain models," Acta Acust. Acust. **88**, 483–492 (2002).
[34]Z. Maekawa, "Noise reduction by screens," Appl. Acoust. **1**, 157–173 (1968).
[35]E. M. Salomons, "Noise barriers in a refracting atmosphere," Appl. Acoust. **47**, 217–238 (1996).
[36]U. J. Kurze, "Noise reduction by barriers," J. Acoust. Soc. Am. **55**, 504–518 (1974).
[37]A. P. Dowling and J. E. Ffowcs Williams, *Sound and Sources of Sound* (Ellis Horwood, Chichester, 1983) pp. 139–144.
[38]E. M. Salomons, "Reduction of the performance of a noise screen due to screen-induced wind-speed gradients. Numerical computations and wind-tunnel experiments," J. Acoust. Soc. Am. **105**, 2287–2293 (1999).
[39]E. M. Salomons, *Computational Atmospheric Acoustics* (Kluwer, Dordrecht, 2001), pp. 289–295.

Real-time calculation of a limiting form of the Renyi entropy applied to detection of subtle changes in scattering architecture

M. S. Hughes[a)]
School of Medicine, Washington University, 660 South Euclid Avenue, Campus Box 8086, St. Louis, Missouri 63110-1093

J. E. McCarthy and M. V. Wickerhauser
Department of Mathematics, Washington University, 660 South Euclid Avenue, Campus Box 8086, St. Louis, Missouri 63110-1093

J. N. Marsh, J. M. Arbeit, R. W. Fuhrhop, K. D. Wallace, T. Thomas, J. Smith, K. Agyem, G. M. Lanza, and S. A. Wickline
School of Medicine, Washington University, 660 South Euclid Avenue, Campus Box 8086, St. Louis, Missouri 63110-1093

(Received 2 April 2009; revised 12 August 2009; accepted 15 August 2009)

Previously a new method for ultrasound signal characterization using entropy H_f was reported, and it was demonstrated that in certain settings, further improvements in signal characterization could be obtained by generalizing to Renyi entropy-based signal characterization $I_f(r)$ with values of r near 2 (specifically $r=1.99$) [M. S. Hughes *et al.*, J. Acoust. Soc. Am. **125**, 3141–3145 (2009)]. It was speculated that further improvements in sensitivity might be realized at the limit $r\rightarrow 2$. At that time, such investigation was not feasible due to excessive computational time required to calculate $I_f(r)$ near this limit. In this paper, an asymptotic expression for the limiting behavior of $I_f(r)$ as $r\rightarrow 2$ is derived and used to present results analogous to those obtained with $I_f(1.99)$. Moreover, the limiting form $I_{f,\infty}$ is computable directly from the experimentally measured waveform $f(t)$ by an algorithm that is suitable for real-time calculation and implementation.
 [DOI: 10.1121/1.3224714]

PACS number(s): 43.60.Bf, 43.60.Lq, 43.60.Cg [EJS] Pages: 2350–2358

I. INTRODUCTION

In an earlier paper[1] we reported on the application of Renyi entropy $I_f(r)$, which is defined for all $r<2$ (r is roughly a reciprocal "temperature"), for the detection of changes in backscattered radio frequency (rf) ultrasound arising from the accumulation of targeted nanoparticles in the neovasculature in the insonified region of a tumor. That study was motivated by the observation that acoustic characterization of sparse collections of targeted perfluorocarbon nanoparticles presented challenges that might require the application of novel types of signal processing.[2] We were able to show that signal processing based on a "moving window" H_f analysis [see Eq. (7)] could detect accumulation of tissue-targeted nanoparticles 30 min following nanoparticle injection. The signal energy, defined as the sum of squares over the same moving window, was unable to distinguish measurements made at any time during the 1 h experiment (as was conventional *B*-mode imaging). Subsequently we determined that moving window $I_f(r)$ analysis, with $r=1.99$, could distinguish the difference in backscatter measured at 0 and 15 min. Reduction in the accumulation time required to reach detectability from 30 to 15 min is clearly of significance: potentially reducing both patient discomfort and increasing clinical throughput. Moreover, although the computational effort to obtain the result precluded its clinical application with currently available equipment, the study raised the possibility of further sensitivity improvements by using values of r closer to the limiting value of 2, where $I_f(r)$ approaches infinity. The purpose of the current study is to investigate the behavior of $I_f(r)$ as $r\rightarrow 2$ by extracting its asymptotic form. While this involves use of the second derivatives of $f(t)$ at its critical points, which can be expected to increase noise in the processing chain output, surprisingly the resulting signal processing scheme does not sacrifice sensitivity. Moreover, the operation count in this approach is lower than that used to produce the signal envelope, which currently is the standard for real-time ultrasonic imaging display, thus demonstrating its suitability for implementation in a real-time imaging system to facilitate detection of molecular epitopes associated with neovasculature in a growing tumor. As our technique is based on moving window analysis of digitized rf, which requires some sacrifice in spatial resolution, clinical implementation of entropy detection would probably follow the same approach currently employed in Doppler "imaging" systems, where the conventional *B*-mode image is color-coded according to the blood cell velocity to present a combined *B*-mode/velocity image; similarly, a *B*-mode/entropy image could be made as well.

[a)]Author to whom correspondence should be addressed. Electronic mail: msh@cmrl.wustl.edu

II. APPROACH

All results in this study were obtained using the density function $w_f(y)$ of the continuous function $y=f(t)$, assumed to underlie the sampled rf data. Subsequently, $w_f(y)$ was used to compute the entropy $I_f(r)$. As described in previous studies $w_f(y)$ corresponds to the density functions used in statistical signal processing.[1] In contradistinction to statistical signal processing, where $f(t)$ is a random function, and often nowhere differentiable, we assume that the noise levels in our apparatus are low enough so that with sufficient signal averaging, noise may be eliminated, or at least reduced to a low enough level, that derivatives of $f(t)$ may be accurately computed. From these derivatives the density function $w_f(y)$ may be computed,[1] which then facilitates calculation of the quantities typically discussed in statistical signal processing (e.g., mean values, variances, and covariances).[3–5] However, in that environment, the density function is usually assumed to be continuous, infinitely differentiable, and to approach zero at infinity. In our case $w_f(y)$ is not so well-behaved and has (integrable) singularities. While this renders calculation of the density function more difficult, applications of entropy imaging based on $w_f(y)$ have shown the cost to be justified in terms of increased sensitivity to subtle changes in scattering architecture that are often undetected by more conventional imaging.

We use the same conventions as in previous studies so that

$$w_f(y) = \sum_{k=1}^{N} |g_k'(y)|, \tag{1}$$

where N is the number of laps [regions of monotonicity of $f(t)$], $g_k(y)$ is the inverse of $f(t)$ in the kth-lap, and if y is not in the range of $f(t)$ in the kth-lap, $g_k'(y)$ is taken to be 0.

We also assume that all experimental waveforms $f(t)$ have a Taylor series expansion valid in the domain [0,1]. Then near a time t_k such that $f'(t_k)=0$,

$$y = f(t) = f(t_k) + \frac{1}{2!} f''(t_k)(t-t_k)^2 + \cdots, \tag{2}$$

where t_k is a lap boundary. On the left side of this point Eq. (2) may be truncated to second order and inverted to obtain

$$g_k(y) \sim t_k \pm \sqrt{2(y-f(t_k))/f''(t_k)}, \tag{3}$$

with

$$|g_k'(y)| \sim 1/\sqrt{2f''(t_k)(y-f(t_k))}. \tag{4}$$

The contribution to $w_f(y)$ from the right side of the lap boundary, from $g_{k+1}(y)$, is the same, so that the overall contribution to $w_f(y)$ coming from the time interval around t_k is

$$|g_k'(y)| \sim \sqrt{2/(f''(t_k)(y-f(t_k)))}, \tag{5}$$

for $0<f(t_k)-y\ll 1$ for a maximum at $f(t_k)$ and $0<y-f(t_k)\ll 1$ for a minimum. Thus, $w_f(y)$ has only a square root singularity (we have assumed that t_k is interior to the interval [0,1]; if not, then the contributions to w_f come from only the left or the right). If, additionally, $f''(t_k)=0$, then the square root singularity in Eq. (4) will become a cube-root singularity, and so on, so that the density functions we consider will have only integrable algebraic singularities.

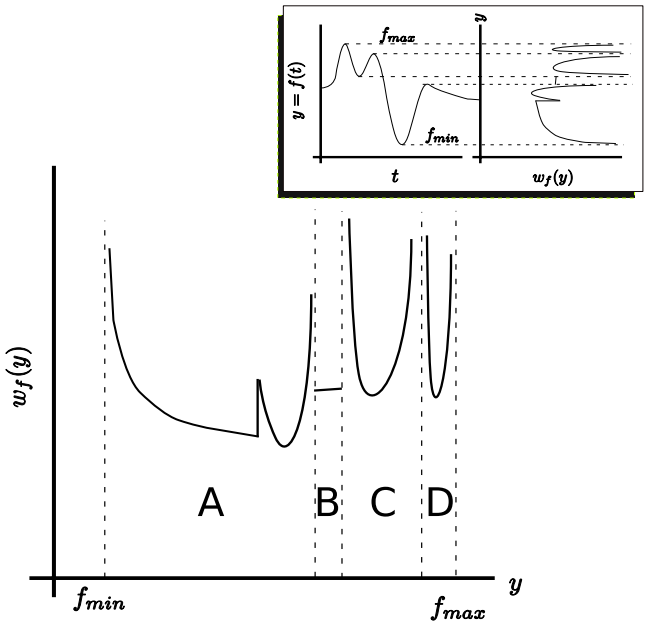


FIG. 1. (Color online) Plot of a typical density function $w_f(y)$ employed in our study. Inset shows a time-domain waveform $f(t)$ with five critical points (left) and their relationship to the singularities of the associated density function $w_f(y)$ (right).

Figure 1 shows a typical waveform (inset) and its associated density function, which may be divided into four different regions, A–D, separated by singularities (dashed lines) corresponding to the critical points of $f(t)$. In general, three types of behavior are possible in $w_f(y)$: continuous, finite jump discontinuity, and integrable singularity. In region A, there are two singularities and a finite jump discontinuity. In region B, $w_f(y)$ is continuous. In C and D, there are singularities at the region boundaries. This figure shows that the density functions possess significantly different attributes from those usually considered in statistical signal processing. To compare the current approach with that usually taken in discussions of "random variable theory," we point out that a "random variable," usually denoted $X(t)$ [instead of our $f(t)$] is nothing more or less than a Lebesgue measurable function. In many applications of random variable theory, $X(t)$ is everywhere continuous but nowhere differentiable (e.g., the Brownian motion), and various means are devised to estimate its probability density function. In the current investigation, it is not necessary to estimate $w_f(y)$ by these means since we may calculate it from $f(t)$. However, as $f(t)$ is assumed to be Lebesgue measurable, it is, in the strictly formal sense, also a random variable, and $w_f(y)$ is its probability density function. As we are investigating a subset of random variables, where it happens that the probability density function may be calculated from the random variable itself, we will carefully refrain from using this term, since it will only raise associations with the reader's mind that are misleading in the current context.

The mathematical characteristics of the singularities of $w_f(y)$ are important in order to guarantee the existence of the following integral on which we base our analysis of signals in this study:

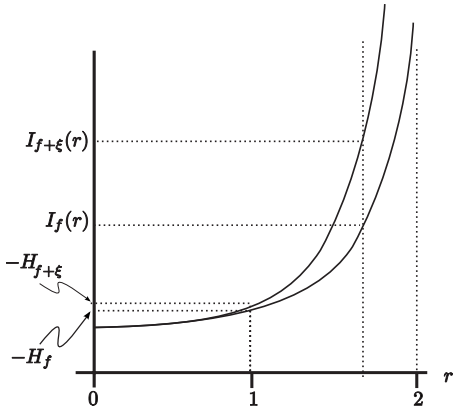


FIG. 2. Plots of $I_f(r)$ and $I_{f+\xi}(r)$ (left) showing that while $I_f(1)=-H_f$ and $I_{f+\xi}(1)=-H_{f+\xi}$ may be close, $I_f(r)$ and $I_{f+\xi}(r)$ diverge as $r\to 2$.

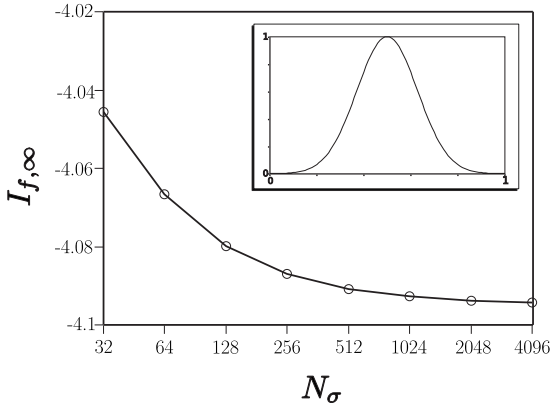


FIG. 3. Simulation for a noise-free Gaussian pulse showing the dependence of $I_{f,\infty}$ on the number of sampled points N_σ.

$$I_f(r)=\frac{1}{1-r}\log\left[\int_{f_{\min}}^{f_{\max}} w_f(y)^r dy\right], \tag{6}$$

which is known as the Renyi entropy.[6] It is similar to the partition function in statistical mechanics with the parameter r playing the role of an "artificial" reciprocal temperature[1,7] (unrelated to the actual physical temperature in the scattering region); moreover, $I_f(r)\to -H_f$, as $r\to 1$, using L'Hôpital's rule, so that I_f is a generalization of H_f as follows:

$$H_f=\int_{f_{\min}}^{f_{\max}} w_f(y)\log w_f(y)dy. \tag{7}$$

Previous studies have shown that this quantity can be more sensitive to subtle changes in scattering architecture than are more commonly used energy-based measures,[2] with subsequent studies demonstrating further sensitivity improvements using I_f at the suitable value of r.[1] For the density functions $w_f(y)$ encountered in our study, $I_f(r)$ is undefined for $r\geq 2$, since as $r\to 2^-$, the integral appearing in Eq. (6) will grow without bound due to the singularities in the density function $w_f(y)$ described in Eq. (5). The behavior as $r\to 2$ is dominated by contributions from these singularities, all of which correspond to critical points of $f(t)$. This behavior is shown in Fig. 2. Moreover, as shown in the figure, it is possible that two slightly different functions, $f(t)$ and $f(t)+\xi(t)$, where ξ is small, may have entropies, H_f and $H_{f+\xi}$, that are close, as shown, but whose Renyi entropies, $I_f(r)$ and $I_{f+\xi}(r)$, diverge as $r\to 2$. Previous studies have shown that this can happen in practice.[1] However, these results left open the possibility of further sensitivity gains. The purpose of the present study is to investigate the possibility of obtaining further sensitivity improvements by pushing toward this limit. As described in the Appendix, the asymptotic form of $I_f(r)$ as $r\to 2$ is given by

$$I_{f,\infty}=\log\left[\sum_{\{t_k|f'(t_k)=0\}}\frac{1}{|f''(t_k)|}\right]. \tag{8}$$

We will use this quantity to generate the images presented in the current study.

III. MATERIALS AND METHODS

A. Numerical computation of $I_{f,\infty}$

Calculation of $I_{f,\infty}$ via Eq. (8) is accomplished by fitting a cubic spline to the experimentally acquired data array using a well-known algorithm, which returns the second derivative of the cubic spline (in an array having the same length as the experimental data) and initializes data structures suitable for rapid computation of its first derivative.[8] Subsequently, an array of corresponding first derivatives is computed and used to bracket the critical points of the spline (i.e., the zero crossings). Linear interpolation is then used to estimate the exact location of the bracketed zero crossings in order to obtain an algorithm suitable for real-time implementation in a medical imaging system. The total operation count is of order N_σ, where N_σ is the number of points processed, and is more than four orders of magnitude faster than the operation count $16\,384N_\sigma$ required to compute $I_f(r)$ used in our previous study.[1] For comparison, we also note that the operation count required to produce the envelope of the same number of points (i.e., to produce a conventional B-mode image) would be of order $N_\sigma\log(N_\sigma)$ since computation of the envelope requires use of the fast Fourier transform; for the value of $N_\sigma=512$ used in our study below, this represents an increase in processing speed of roughly ninefold.

B. Simulations

The convergence properties, stability in the presence of noise, and effects of quantization error and sampling rate have been extensively evaluated using simulated data. Several types of waveforms have been investigated: Gaussian and parabolic waveforms, for which the exact value of $I_{f,\infty}$ may be computed and linear combinations of exponentially damped sine waves that qualitatively resemble backscattered ultrasonic waveforms. Several carefully chosen example simulations illustrate guidelines for application of our algorithm in order to avoid potential artifacts produced by experimental factors.

The first of these is Fig. 3, which shows a plot of $I_{f,\infty}$ for a noise-free Gaussian pulse $f(t)=e^{-30(t-0.5)^2}$ for values of N_σ ranging from 32, 64, 128,…, 8192. Even at $N_\sigma=32$ the estimated value of $I_{f,\infty}$ is within 1% of the exact value of

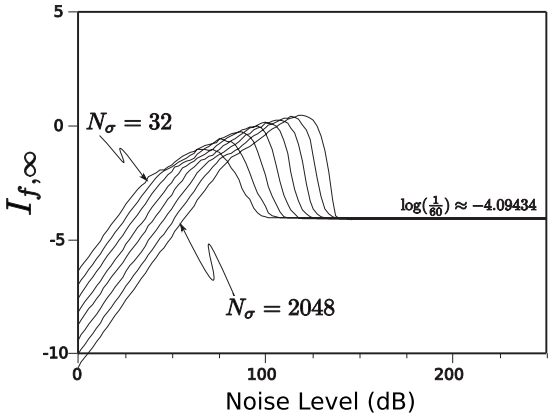


FIG. 4. Simulation for a Gaussian pulse showing the dependence of $I_{f,\infty}$ on the number of sampled points N_σ and noise level.

$\log[1/60] = -4.094$. For moving window analysis of experimental data, N_σ is the length of the moving window. Choosing its length requires making trade-offs between sensitivity (smaller N_σ implying loss of sensitivity, but increased spatial resolution), noise level (smaller N_σ implying increased noise, but increased spatial resolution), and spatial resolution.

However, noise can have a significant effect on the calculation of $I_{f,\infty}$. Figure 4 illustrates the impact of noise on the Gaussian pulse ($f(t)=e^{-30(t-0.5)^2}$) that was just discussed. As N_σ ranges from 32, 64, 128,…, 8192 and noise levels ranges from 0 to 150 dB, the calculated value of $I_{f,\infty}$ can vary by over 100% of its actual value. Eventually, as N_σ increases and the noise level drops, our algorithm converges to a stable value. However, as the plots indicate, the noise requirements for a single peak function such as the Gaussian peak are quite stringent, being greater than 100 dB to obtain 10% accuracy.

These requirements are less stringent if $f(t)$ has several critical points. An example is shown in Fig. 5, which plots $I_{f,\infty}$ for values of N_σ ranging from 32, 64, 128,…, 2048, and for noise levels ranging from 0 to 150 dB for the Gaussian modulated pulse $f(t)=e^{-10(t-0.5)^2}\sin(20\pi(t-0.5))+0.7\sin(20\pi(t-0.5))+0.7\sin(10\pi(t-0.5))$. As the plots indicate the noise requirements for a multipeak peak waveform $f(t)$ are far less stringent with 87% accuracy being obtained at about 20 dB noise level for $N_\sigma=512$ (plotted

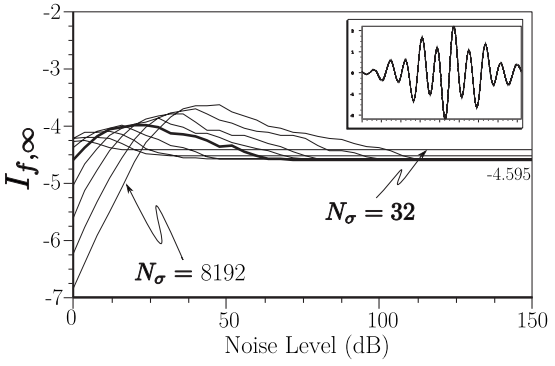


FIG. 5. Simulation for an unquantized Gaussian modulated pulse showing the dependence on the number of sampled points N_σ and noise level.

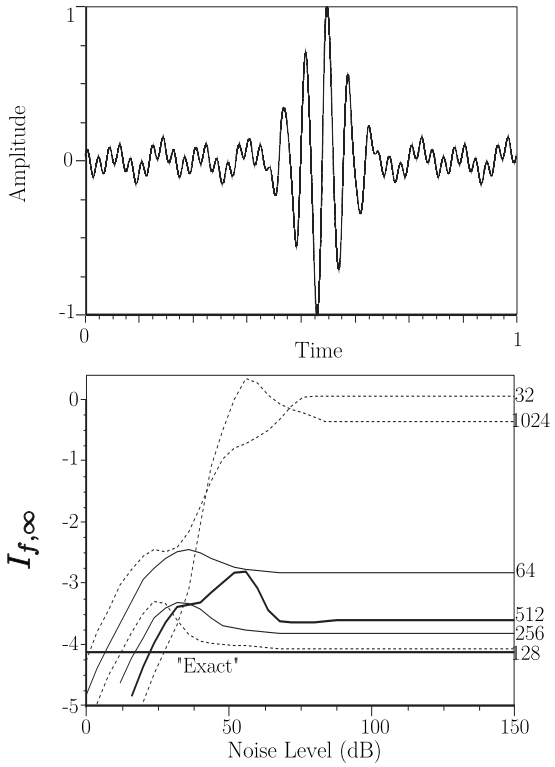


FIG. 6. Top panel: the simulated backscatter signal described in the text. Bottom panel: plot showing the dependence of $I_{f,\infty}$ on the number of sampled points N_σ and noise level at 8-bit quantization. The heavy black line labeled "Exact" is at 4.149, the limiting value of $I_{f,\infty}$ obtained from our algorithm in the unquantized, noise-free case with $N_\sigma=8192$.

using a heavier line in the plot family since these parameters match values used in the experimental portion of our study).

Figure 6 shows a plot of $I_{f,\infty}$ for values of N_σ ranging from 32, 64, 128,…, 8192 for noise levels ranging from 0 to 150 dB for the simulated pulse $f(t)=e^{-150(t-0.55)^2}\times\sin(40\pi(t-0.55))+0.7\sin(80\pi(t-0.55))+0.7\sin(20\pi(t-0.55))+0.03\sin(10\pi(t-0.55))$. The "exact" answer is $-4.149\,073$, found by running our algorithm with noise level set to zero, no quantization error, and $N_\sigma=8192$, and is also shown in the plot. The corresponding values of N_σ are indicated on the right side of the figure. For values of $N_\sigma \leq 512$ the error is less than 13%. We also note that for larger values of N_σ and lower levels of noise, our algorithm diverges with $I_{f,\infty}$ becoming large and positive. This occurs only in quantized simulations and is the result of the long perfectly flat segments in the quantized data. This is an easily detected fault and, since the $I_{f,\infty}$ images used in our experimental study had pixel values of approximately 7 bits/symbol in magnitude on the regions used to estimate accumulation of targeted nanoparticles, can be ruled out as a possible artifact in our study.

C. Nanoparticles for molecular imaging

A cross-section of the spherical liquid nanoparticles used in our study is diagramed in Fig. 7. For *in vivo* imaging we formulated nanoparticles targeted to $\alpha_v\beta_3$-integrins of neovascularity in cancer by incorporating an "Arg-Gly-Asp" mimetic binding ligand into the lipid layer. Methods devel-

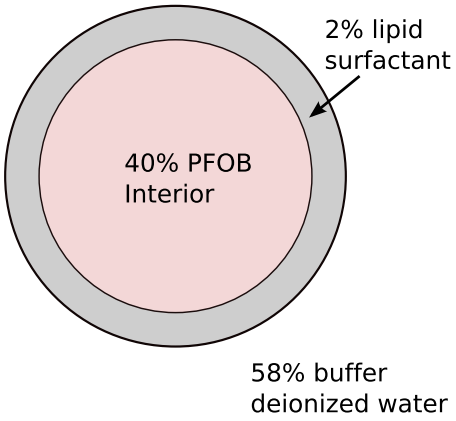


FIG. 7. (Color online) A cross-sectional diagram of the nanoparticles used in our study.

oped in our laboratories were used to prepare perfluorocarbon (perfluorooctylbromide, which remains in a liquid state at body temperature and at the acoustic pressures used in this study[9]) emulsions encapsulated by a lipid-surfactant monolayer.[10,11] The nominal sizes for each formulation were measured with a submicron particle analyzer (Malvern Zetasizer, Malvern Instruments). Particle diameter was measured at 200 ± 30 nm.

D. Animal model

The study was performed according to an approved animal protocol and in compliance with guidelines of the Washington University Institutional Animal Care and Use Committee.

The model used is the transgenic K14-HPV16 mouse in which the ears typically exhibit squamous metaplasia, a precancerous condition, associated with abundant neovasculature that expresses the $\alpha_v\beta_3$-integrin. Eight of these transgenic mice[12,13] were treated with 1.0 mg/kg intravenous of either $\alpha_v\beta_3$-targeted nanoparticles ($n=4$) or untargeted nanoparticles ($n=4$) and imaged dynamically for 1 h using a research ultrasound imager (Vevo 660 40 MHz probe) modified to store digitized rf waveforms acquired at 0, 15, 30, and 60 min after injection of nanoparticles. In both targeted and untargeted cases, the mouse was placed on a heated platform maintained at 37 °C, and anesthesia was administered continuously with isoflurane gas (0.5%).

E. Ultrasonic data acquisition

A diagram of our apparatus is shown in Fig. 8. rf data were acquired with a research ultrasound system (Vevo 660, Visualsonics, Toronto, Canada), with an analog port and a sync port to permit digitization. The tumor was imaged with a 40 MHz single element "wobbler" probe and the rf data corresponding to single frames were stored on a hard disk for later off-line analysis. The frames (acquired at a rate of 40 Hz) consisted of 384 lines of 4096 8-bit words acquired at a sampling rate of 500 MHz using a Gage CS82G digitizer card (connected to the analog-out and sync ports of the Vevo) in a controller PC. Each frame corresponds spatially to a region 0.8 cm wide and 0.3 cm deep.

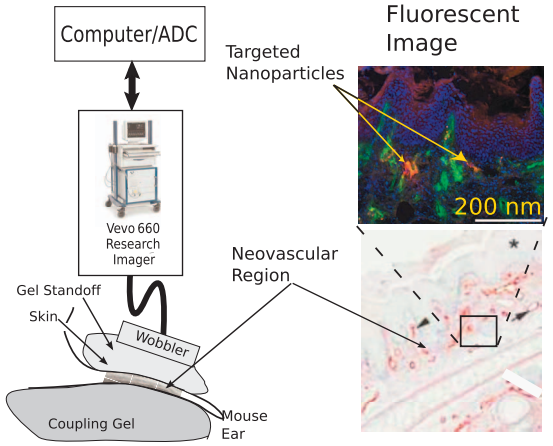


FIG. 8. A diagram of the apparatus used to acquire rf data backscattered from K14-HPV16 transgenic mouse ears *in vivo* together with a histologically stained section of the ear indicating portions where $\alpha_v\beta_3$-targeted nanoparticles could adhere and a fluorescent image demonstrating presence of targeted nanoparticles.

The wobbler transducer used in this study is highly focused (3 mm in diameter) with a focal length of 6 mm and a theoretical spot size of 80×1100 μm (lateral beam width $\times$ depth of field at -6 dB), so that the imager is most sensitive to changes occurring in the region swept out by the focal zone as the transducer is "wobbled." Accordingly, a gel standoff was used, as shown in Fig. 8, so that this region would contain the mouse ear.

A close-up view showing the placement of transducer, gel standoff, and mouse ear is shown in the bottom of the figure. Superposed on the diagram is a *B*-mode gray scale image (i.e., logarithm of the analytic signal magnitude). Labels indicate the location of skin (top of image insert), the structural cartilage in the middle of the ear, and a short distance below this, the echo from the skin at the bottom of the ear. Directly above this is an image of a histological specimen extracted from a K14-HPV16 transgenic mouse model that has been magnified 20 times to permit better assessment of the thickness and architecture of the sites where $\alpha_v\beta_3$-targeted nanoparticle might attach (red by β_3 staining). Skin and tumor are both visible in the image. On either side of the cartilage (center band in image), extending to the dermal-epidermal junction, is the stroma. It is filled with neoangiogenic microvessels. These microvessels are also decorated with $\alpha_v\beta_3$ nanoparticles as indicated by the fluorescent image (labeled, in the upper right of the figure) of a bisected ear from an $\alpha_v\beta_3$-injected K14-HPV16 transgenic mouse. It is in this region that the $\alpha_v\beta_3$-targeted nanoparticles are expected to accumulate, as indicated by the presence of red β_3 stain in the magnified image of a histological specimen also shown in the image.

F. Ultrasonic data processing

Each of the 384 rf lines in the data was first up-sampled from 4096 to 8192 points, using a cubic spline fit to the original data set in order to improve the stability of the thermodynamic receiver algorithms. A by-product of this "order N_σ" algorithm is simultaneous output of a corresponding array of second derivative values of the fit function.[8] Next, a

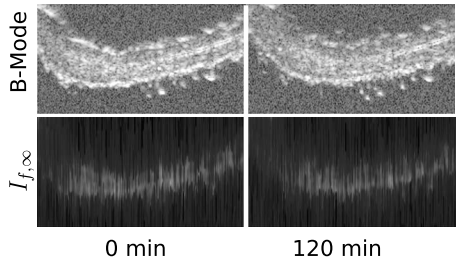


FIG. 9. Top row: conventional *B*-mode images at 0 min (left), and after injection of $\alpha_v\beta_3$-targeted nanoparticles (right). Bottom row: corresponding $I_{f,\infty}$ images.

moving window analysis was performed on the second derivative data set, using Eq. (8) to compute $I_{f,\infty}$, by moving a rectangular window (512 points long, 0.512 μs) in 0.064 μs steps (64 points), resulting in 121 window positions within the output data set. This produced an image for each time point in the experiment. The window length was chosen to match that used in previous studies;[1,2] it corresponds to the heavy black curve shown in Fig. 6. Analyses were also performed using window lengths of 256 (0.256 μs) and 128 points (0.256 μs). While they also produced statistically significant changes in $I_{f,\infty}$ versus time, post-injection, the resulting $I_{f,\infty}$ versus time curves were noisier, and required 1 h, post-injection, to exhibit statistically significant changes. As discussed previously, the optimum choice of window length requires trade-offs between sensitivity, noise level, and spatial resolution. In Sec. IV we discuss the 512 point moving window length results since they correspond most closely to previous results, which were supported by independent histological results,[1,2] and produced images with sufficient spatial resolution to indentify relevant anatomical features in the mouse ear. Additionally, a major goal of this study was to assess the numerical stability of the algorithm, which is based on the second derivative of an experimentally measured data set, and thus contaminated by noise. Ordinarily, estimation of just the first derivative is difficult. However, in our application, the effects of noise might be mitigated by two factors: The second derivative is obtained from a global fit to the data, and it appears in the denominator of the expression for receiver output so that values having large error are likely to make small contributions to the sum appearing in Eq. (8).

G. Image processing

All rf data were processed off-line to reconstruct $I_{f,\infty}$ images. Total analysis time using the new algorithm was less than 5 min on an eight core desktop computer [compared to the roughly week-long time required to execute the $I_f(1.99)$ analysis on a cluster of just over 20 computers that was reported previously[1]]. A representative set of these images is shown in the bottom row of Fig. 9. For comparison, the top row shows conventional *B*-mode images, i.e., logarithm of the signal envelope. The left columns show images constructed from the rf data sets acquired 0 min after injection while the right column shows the images constructed from data acquired 120 min post-injection. The look-up-table of the entropy images have been inverted to produce a display

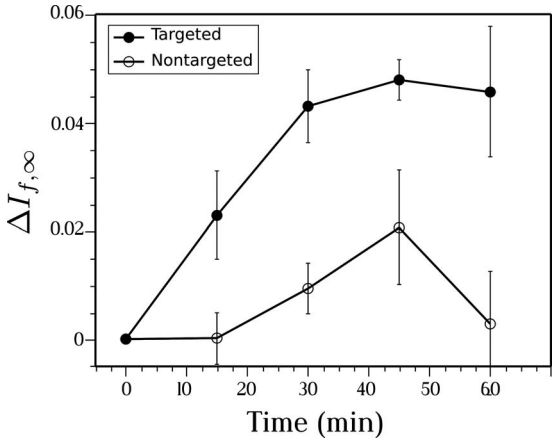


FIG. 10. $I_{f,\infty}$ image enhancement, i.e., change relative to 0 min, obtained after injection of $\alpha_v\beta_3$-targeted nanoparticles (closed circles) and nontargeted nanoparticles (open circles) into four K14-HPV16 transgenic mice in each case.

in which pixels corresponding to tissue are brighter than surrounding pixels, in order to facilitate comparison with the conventional images. As expected, the conventional images exhibit higher spatial resolution compared with the $I_{f,\infty}$ images, which employ a moving window in their construction. While the window is discernible in the entropy images, it is not a cause of major concern since our goal is automatic quantitative detection of changes in scattering architecture in the physical region represented by the image. As may be seen from the figure, the bright regions in $I_{f,\infty}$ and conventional images are correlated. Moreover, as time passes from 0 to 120 min, both rows of images show a reduction in contrast between tissue of the ear (bright regions) and darker background (corresponding to the gel couplant). In the $I_{f,\infty}$ images this corresponds to an increase in $I_{f,\infty}$. As discussed below this effect is consistently observed in all $I_{f,\infty}$ images from the group of K14-HPV16 transgenic mice injected with $\alpha_v\beta_3$-targeted nanoparticles, and it is not observed in $I_{f,\infty}$ images from any of the control groups. As stated in previous publications, there are no statistically significant changes in signal energy images in any of the groups studied.[2]

Subsequently, a histogram of pixel values for the composite of the 0, 15, 30, and 60 min images was computed as described in previous papers.[1,2] Image segmentation of each type of image, at each time point in the experiment, was then performed automatically using its corresponding histogram according to the following threshold criterion: The lowest 7% of pixel values were classified as "targeted" tissue, while the remaining were classified as "untargeted" (histogram analysis was also performed using 90% and 87% thresholds, with 93% having the best statistical separation between time points). The mean value of pixels classified as targeted was computed at each time post-injection.

IV. RESULTS AND DISCUSSION

The results obtained after injection of targeted nanoparticles and nontargeted nanoparticles by $I_{f,\infty}$ receiver are shown in Fig. 10. Both curves show the time evolution of the change (relative to 0 min) in mean value of receiver output in the enhanced regions of images obtained from the four ani-

mals in the targeted and the four animals in the nontargeted groups. Standard error bars are shown with each point. At 15 min the change in mean value of $I_{f,\infty}$ is more than two standard errors from zero, implying statistical significance at the 95% level. There is no statistically significant change in image brightness for the nontargeted nanoparticles' group. As the results show, the algorithm for computation of $I_{f,\infty}$ is stable in the presence of experimental noise.

The results presented in this paper extend earlier studies where it was shown that an entropy-based measure H_f was able to detect targeted nanoparticles in tumor neovasculature[2] after 30 min of accumulation time. Subsequently, the time required to detect targeted nanoparticles was reduced to 15 min using a generalization of entropy $I_f(r)$, with $r=1.99$, although the time required for signal analysis was greatly increased.[1] In the current study based on $I_{f,\infty}$, the analysis time has been reduced from days to minutes using an algorithm suitable for real-time implementation, while maintaining sensitivity that permits detection of nanoparticle accumulation at 15 min.

Real-time performance appears to have been purchased at the price of reduced statistical sensitivity, in view of prior observation that $I_f(1.99)$ separated by over five standard errors from 0 at 15 min (Ref. 1) as compared to the two standard error separation obtained with the real-time receiver (see Fig. 10). It is possible that preprocessing of the data by bandpass filtering might improve the statistical performance of the algorithm without significant increase in computational overhead. This will be studied in a future report.

ACKNOWLEDGMENTS

This study was funded by NIH Grant Nos. EB002168, HL042950, and CO-27031, and NSF Grant No. DMS 0501079. The research was carried out at the Washington University Department of Mathematics and the School of Medicine.

APPENDIX: DERIVATION OF ASYMPTOTIC FORM

As described in Sec. II, the limiting form of $I_f(r)$ as $r\to 2$ will now be derived. The first step is to observe that the integral in Eq. (6) may split into two parts, one corresponding to the region where the function is clearly bounded and one corresponding to its singularities as shown in Fig. 11. Thus,

$$\begin{aligned}\int_{f(t_k)}^{f(t_{k+1})-\delta_{k+1}} w_f(y)^{2-\epsilon}dy &= \int_{f(t_k)}^{f(t_k)+\delta_k} w_f(y)^{2-\epsilon}dy + \int_{f(t_k)+\delta_k}^{f(t_{k+1})-\delta_{k+1}} w_f(y)^{2-\epsilon}dy \\ &= \int_{f(t_k)}^{f(t_k)+\delta_k} w_f(y)^{2-\epsilon}dy + B_k, \end{aligned} \tag{A1}$$

where we have written B_k for the integral over the unshaded region between $f(t_k)+\delta_k$ and $f(t_{k+1})-\delta_{k+1}$ in Fig. 11. We observe that B_k is bounded as $\epsilon\to 0$, while the integral appearing in Eq. (A1) is not.

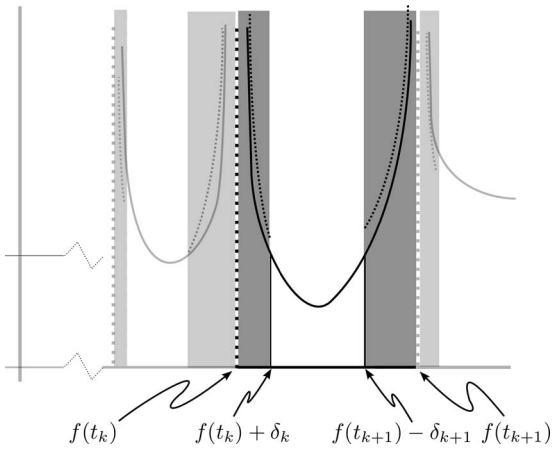


FIG. 11. An enlarged plot of a singularity of the density function $w_f(y)^{2-\epsilon}$ (solid curve) and Eq. (5) (dashed curves); quantities relevant for derivation of Eq. (A12). As the shaded regions shrink the ratio between the dashed and solid curves approaches 1. The darker shading corresponds to the region discussed in the text.

Next, we consider the small interval of length δ_k near the singularity of $w_f(f(t_k))$ (shaded regions of Fig. 11). This is the singularity corresponding to the kth extrema of $f(t)$: $f(t_k)$; also shown is the adjacent singularity corresponding to an extrema of $f(t)$ at t_{k+1}. The dashed lines in these regions represent the one over square root limiting form described in Eq. (5). By choosing δ_k small enough we may make the ratio of the solid and dashed curves arbitrarily close to 1. In other words, Eqs. (1) and (5) tell us that in these shaded regions the following difference can be made as small as we like:

$$\left| w_f(y-\delta_k)/\frac{a_k}{\sqrt{y-f(t_k)}} - 1 \right|, \tag{A2}$$

where $a_k=\sqrt{2/f''(t_k)}=\sqrt{2/|f''(t_k)|}$ [assuming a minimum at $f(t_k)$, the argument for a maximum is similar]. Moreover, if a particular choice of δ_k yields the desired accuracy, i.e., makes the difference small enough, choosing a smaller value of δ_k will produce greater accuracy. Since the number of extrema in our time-domain function $f(t)$ is finite, we pick the minimum δ_k, call it δ, yielding the desired accuracy in all of the shaded regions [i.e., at all singular points of $w_f(y)$]. With this choice of δ Eq. (A1) becomes

$$\int_{f(t_k)}^{f(t_{k+1})-\delta} w_f(y)^{2-\epsilon}dy = \int_{f(t_k)}^{f(t_k)+\delta} w_f(y)^{2-\epsilon}dy + \tilde{B}_k, \tag{A3}$$

and Eq. (A2) becomes

$$\left| w_f(y-\delta)/\frac{a_k}{\sqrt{y-f(t_k)}} - 1 \right| < E, \tag{A4}$$

or

$$w_f(y-\delta)/a_k/\sqrt{y-f(t_k)} = 1 \pm E(y), \tag{A5}$$

where $E>|E(y)|$ may be chosen to be as small as we like by choosing small enough δ. As a result

$$w_f(y-\delta)^{2-\epsilon} = \left(\frac{a_k}{\sqrt{y-f(t_k)}}\right)^{2-\epsilon}[1 \pm E(y)]^{2-\epsilon}$$

$$= \left(\frac{a_k}{\sqrt{y-f(t_k)}}\right)^{2-\epsilon}[1 \pm \tilde{E}(y)], \tag{A6}$$

where, once again, $\tilde{E}(y)$ may be made arbitrarily small, i.e., for every $\tilde{E}>0$ there exists some $\delta>0$ such that $\tilde{E}>|\tilde{E}(y)|$ for all y in between $f(t_k)$ and $f(t_k)+\delta$.

Combining Eqs. (A3) and (A6) now yields

$$\int_{f(t_k)}^{f(t_k)+\delta} w_f(y)^{2-\epsilon}dy + \tilde{B}_k$$

$$= \int_{f(t_k)}^{f(t_k)+\delta}\left(\frac{a_k}{\sqrt{y-f(t_k)}}\right)^{2-\epsilon}[1 \pm \tilde{E}(y)]dy + \tilde{B}_k$$

$$= \int_{f(t_k)}^{f(t_k)+\delta}\left(\frac{a_k}{\sqrt{y-f(t_k)}}\right)^{2-\epsilon}dy$$

$$\pm \int_{f(t_k)}^{f(t_k)+\delta}\left(\frac{a_k}{\sqrt{y-f(t_k)}}\right)^{2-\epsilon}\tilde{E}(y)dy + \tilde{\tilde{B}}. \tag{A7}$$

The second integral above may be bounded by

$$\left|\int_{f(t_k)}^{f(t_k)+\delta}\left(\frac{a_k}{\sqrt{y-f(t_k)}}\right)^{2-\epsilon}\tilde{E}(y)dy\right|$$

$$\leq \int_{f(t_k)}^{f(t_k)+\delta}\left(\frac{a_k}{\sqrt{y-f(t_k)}}\right)^{2-\epsilon}|\tilde{E}(y)|dy$$

$$\leq \int_{f(t_k)}^{f(t_k)+\delta}\left(\frac{a_k}{\sqrt{y-f(t_k)}}\right)^{2-\epsilon}\tilde{E}dy$$

$$\leq \tilde{E}\int_{f(t_k)}^{f(t_k)+\delta}\left(\frac{a_k}{\sqrt{y-f(t_k)}}\right)^{2-\epsilon}dy. \tag{A8}$$

This inequality may be converted to an equality by replacing the $\tilde{E}$ factor with a smaller (positive) number. In general, this number will depend on the behavior of $w_f(y)$ near the singular point $y=f(t_k)$. For clarity, we denote this constant by $\tilde{E}_k$. With this notation, Eq. (A8) becomes

$$\int_{f(t_k)}^{f(t_k)+\delta}\left(\frac{a_k}{\sqrt{y-f(t_k)}}\right)^{2-\epsilon}\tilde{E}(y)dy$$

$$= \tilde{E}_k\int_{f(t_k)}^{f(t_k)+\delta}\left(\frac{a_k}{\sqrt{y-f(t_k)}}\right)^{2-\epsilon}dy, \tag{A9}$$

where $\tilde{E}\geq\tilde{E}_k>0$ and hence may also be made as small as we wish by reducing δ. The common integral appearing in Eqs. (A7) and (A8) may be computed as

$$\int_{f(t_k)}^{f(t_k)+\delta}\left(\frac{a_k}{\sqrt{y-f(t_k)}}\right)^{2-\epsilon}dy = a_k^{2-\epsilon}\int_{f(t_k)}^{f(t_k)+\delta}(y-f(t_k))^{1-\epsilon/2}dy$$

$$= a_k^{2-\epsilon}\left.\frac{(y-f(t_k))^{\epsilon/2}}{\epsilon/2}\right|_{f(t_k)}^{f(t_k)+\delta}$$

$$= a_k^{2-\epsilon}\frac{(f(t_k)+\delta-f(t_k))^{\epsilon/2}}{\epsilon/2}$$

$$= \frac{2a_k^2\delta^{\epsilon/2}}{\epsilon}, \tag{A10}$$

so that Eq. (A7) becomes

$$\int_{f(t_k)}^{f(t_{k+1})-\delta} w_f(y)^{2-\epsilon}dy = \frac{2a_k^2\delta^{\epsilon/2}}{\epsilon} \pm \tilde{E}_k\frac{2a_k^2\delta^{\epsilon/2}}{\epsilon} + \tilde{\tilde{B}}_k$$

$$= \frac{2a_k^2\delta^{\epsilon/2}}{\epsilon}[1 \pm \tilde{E}_k] + \tilde{\tilde{B}}_k, \tag{A11}$$

which we sum over all minima to obtain

$$= \sum_{\substack{k \\ f''(t_k)>0}} \frac{2a_k^2\delta^{\epsilon/2}}{\epsilon}[1 \pm \tilde{E}_k] + \tilde{\tilde{B}}_k, \tag{A12}$$

a sum of bounded and unbounded terms, whose unbounded term is computable directly from the experimentally accessible function $f(t)$ using $a_k=\sqrt{2/f''(t_k)}=\sqrt{2/|f''(t_k)|}$.

For the maximum we have the asymptotic term

$$\int_{f(t_k)-\delta}^{f(t_k)}\left(\frac{a_k}{\sqrt{f(t_k)-y}}\right)^{2-\epsilon}dy. \tag{A13}$$

So that the contribution to Eq. (6) from all of the maxima becomes

$$\int_{f(t_k)-\delta}^{f(t_k)}\left(\frac{a_k}{\sqrt{f(t_k)-y}}\right)^{2-\epsilon}dy = a_k^{2-\epsilon}\int_{f(t_k)-\delta}^{f(t_k)}(f(t_k)-y)^{1-\epsilon/2}dy$$

$$= a_k^{2-\epsilon}\left.\frac{(f(t_k)-y)^{\epsilon/2}}{\epsilon/2}\right|_{f(t_k)-\delta}^{f(t_k)}$$

$$= a_k^{2-\epsilon}\frac{(f(t_k)-f(t_k)+\delta)^{\epsilon/2}}{\epsilon/2}$$

$$= \frac{2a_k^2\delta^{\epsilon/2}}{\epsilon}, \tag{A14}$$

we now have a different expression for $a_k=\sqrt{-2/f''(t_k)}$ $=\sqrt{2/|f''(t_k)|}$.

Adding the contributions for the maxima and minima we obtain

$$\int_{f_{\min}}^{f_{\max}} w_f(y)^{2-\epsilon}dy = \sum_{\{t_k|f'(t_k)=0\}} \frac{2a_k^2\delta^{\epsilon/2}}{\int}[1 \pm \tilde{E}_k] + \tilde{\tilde{B}}_k$$

$$= \sum_{\{t_k|f'(t_k)=0\}} \frac{4\delta^{\epsilon/2}}{\epsilon|f''(t_k)|}[1 \pm \tilde{E}_k] + \tilde{\tilde{B}}_k. \tag{A15}$$

Cross multiplying by ϵ

$$\epsilon \int_{f_{\min}}^{f_{\max}} w_f(y)^{2-\epsilon} dy = \sum_{\{t_k | f'(t_k)=0\}} \frac{4\delta^{\epsilon/2}}{|f''(t_k)|}[1 \pm \tilde{E}_k] + \epsilon \tilde{\tilde{B}}_k, \tag{A16}$$

taking the logarithm of both sides and letting $\epsilon \to 0$ we have

$$\lim_{\epsilon \to 0} \left(\log \int + \log \left[\int_{f_{\min}}^{f_{\max}} w_f(y)^{2-\int} dy \right] \right) = \log \left[4 \sum_{\{t_k | f'(t_k)=0\}} \frac{1}{|f''(t_k)|}[1 \pm \tilde{E}_k] \right]. \tag{A17}$$

Now taking the limit $\delta \to 0$ so that the $\tilde{E}_k \to 0$ we obtain

$$\lim_{\int \to 0} \left(\log \epsilon + \log \left[\int_{f_{\min}}^{f_{\max}} w_f(y)^{2-\epsilon} dy \right] \right) = \log \left[4 \sum_{\{t_k | f'(t_k)=0\}} \frac{1}{|f''(t_k)|} \right]. \tag{A18}$$

This shows that as $\epsilon \to 0$, the leading term in $\log \int w_f(y)^{2-\epsilon} dy$ always behaves like $\log 1/\epsilon$, regardless of $f(t)$; but the next term in the asymptotic expansion, the right-hand side of Eq. (A18), does depend critically on $f(t)$, and is the quantity we seek.

Multiplying both sides by $1/(1-r) = 1/(1-2+\epsilon) \to -1$ and then cancelling minus signs on both sides of the equation, we obtain

$$\lim_{\epsilon \to 0} (-\log \epsilon - I_f(2-\epsilon)) = -\log \left[4 \sum_{\{t_k | f'(t_k)=0\}} \frac{1}{|f''(t_k)|} \right]. \tag{A19}$$

For imaging applications, where offset removal and rescaling are typically performed when pixel values are assigned, we define the new quantity

$$I_{f,\infty} \equiv -\lim_{\epsilon \to 0} I_f(2-\epsilon) - \log 4 + \log \epsilon = \log \left[\sum_{\{t_k | f'(t_k)=0\}} \frac{1}{|f''(t_k)|} \right]. \tag{A20}$$

We will use this quantity to generate the images presented in Sec. IV.

[1] M. S. Hughes, J. E. McCarthy, J. N. Marsh, J. M. Arbeit, R. G. Neumann, R. W. Fuhrhop, K. D. Wallace, T. Thomas, J. Smith, K. Agyem, D. R. Znidersic, B. N. Maurizi, S. L. Baldwin, G. M. Lanza, and S. A. Wickline, "Application of Renyi entropy to detect subtle changes in scattering architecture," J. Acoust. Soc. Am. **125**, 3141–3145 (2009).

[2] M. S. Hughes, J. E. McCarthy, J. N. Marsh, J. M. Arbeit, R. G. Neumann, R. W. Fuhrhop, K. D. Wallace, D. R. Znidersic, B. N. Maurizi, S. L. Baldwin, G. M. Lanza, and S. A. Wickline, "Properties of an entropy-based signal receiver with an application to ultrasonic molecular imaging," J. Acoust. Soc. Am. **121**, 3542–3557 (2007).

[3] R. S. Bucy and P. D. Joseph, *Filtering for Stochastic Processes With Applications to Guidance* (Chelsea, New York, 1987).

[4] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications* (MIT, Cambridge, MA, 1949).

[5] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series* (Chelsea, New York, 1984).

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley-Interscience, New York, 1991).

[7] R. Tolman, *The Principles of Statistical Mechanics* (Dover, New York, 1979).

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

[9] M. Hughes, J. Marsh, C. Hall, R. W. Fuhrhop, E. K. Lacy, G. M. Lanza, and S. A. Wickline, "Acoustic characterization in whole blood and plasma of site-targeted nanoparticle ultrasound contrast agent for molecular imaging," J. Acoust. Soc. Am. **117**, 964–72 (2005).

[10] S. Flacke, S. Fischer, M. J. Scott, R. J. Fuhrhop, J. S. Allen, M. McLean, P. Winter, G. A. Sicard, P. J. Gaffney, S. A. Wickline, and G. M. Lanza, "Novel MRI contrast agent for molecular imaging of fibrin implications for detecting vulnerable plaques," Circulation **104**, 1280–1285 (2001).

[11] G. M. Lanza, K. D. Wallace, M. J. Scott, W. P. Cacheris, D. R. Abendschein, D. H. Christy, A. M. Sharkey, J. G. Miller, P. J. Gaffney, and S. A. Wickline, "A novel site-targeted ultrasonic contrast agent with broad biomedical application," Circulation **94**, 3334–3340 (1996).

[12] J. M. Arbeit, R. R. Riley, B. Huey, C. Porter, G. Kelloff, R. Lubet, J. M. Ward, and D. Pinkel, "Chemoprevention of epidermal carcinogenesis in k14-hpv16 transgenic mice," Cancer Res. **59**, 3610–3620 (1999).

[13] J. M. Arbeit, K. Mnger, P. M. Howley, and D. Hanahan, "Progressive squamous epithelial neoplasia in k14-human papillomavirus type 16 transgenic mice," J. Virol. **68**, 4358–4368 (1994).

Effect of reflected and refracted signals on coherent underwater acoustic communication: Results from the Kauai experiment (KauaiEx 2003)

Daniel Rouseff
Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105

Mohsen Badiey and Aijun Song
College of Earth, Ocean, and Environment, University of Delaware, Newark, Delaware 19716

(Received 26 September 2008; revised 30 July 2009; accepted 31 July 2009)

The performance of a communications equalizer is quantified in terms of the number of acoustic paths that are treated as usable signal. The analysis uses acoustical and oceanographic data collected off the Hawaiian Island of Kauai. Communication signals were measured on an eight-element vertical array at two different ranges, 1 and 2 km, and processed using an equalizer based on passive time-reversal signal processing. By estimating the Rayleigh parameter, it is shown that all paths reflected by the sea surface at both ranges undergo incoherent scattering. It is demonstrated that some of these incoherently scattered paths are still useful for coherent communications. At range of 1 km, optimal communications performance is achieved when six acoustic paths are retained and all paths with more than one reflection off the sea surface are rejected. Consistent with a model that ignores loss from near-surface bubbles, the performance improves by approximately 1.8 dB when increasing the number of retained paths from four to six. The four-path results though are more stable and require less frequent channel estimation. At range of 2 km, ray refraction is observed and communications performance is optimal when some paths with two sea-surface reflections are retained. [DOI: 10.1121/1.3212925]

PACS number(s): 43.60.Dh, 43.60.Tj, 43.30.Re, 43.60.Gk [DRD] Pages: 2359–2366

I. INTRODUCTION

Coherent underwater acoustic communication is made difficult by the multipath spread and temporal variability typical of propagation in the ocean. An equalizer attempts to compensate for the intersymbol interference introduced by these effects. For a communications engineer designing an equalizer, the scattering function is a convenient way to model the acoustic telemetry channel.[1] The scattering function describes how transmitted signal power gets spread both in time and in frequency. Both types of signal spread are relevant to equalizer design: The time spread extent determines how many taps are needed by the equalizer while the frequency spread determines how often the taps must be updated if the equalizer is to adapt to the time-varying environment. Designing a practical communications system means making inevitable compromises. To keep the number of taps and the update rates manageable, for example, only a few dominant acoustic paths might be retained as usable signal while other paths are neglected and treated as noise.

The present paper is a case study where the design parameters for an equalizer are varied and the resulting effect on communications performance then quantified using real data. An important feature of the study is that the acoustic telemetry data are augmented by detailed environmental measurements that characterize the communications channel. The optimal values for the equalizer design parameters are thereby related to the number of acoustic paths retained in the processing, the level of scattering they undergo, and the rate at which they fluctuate. Both the refraction of ray paths from the depth-dependent sound speed and the reflection of ray paths from the time-varying sea surface are shown to be important. Paths that undergo incoherent scattering by the rough sea surface are shown to still be useful for coherent acoustic communications under some circumstances. Tradeoffs between processor performance and complexity, in terms of the number of retained paths and environmental variability, are quantified. The sensitivity of the optimal design parameters to the range between the source and receiver is also studied.

The equalizer used in the present case study is outlined in Sec. II. The equalizer is based on passive time-reversal signal processing first proposed for acoustic communications by Dowling[2] and tested in subsequent field experiments.[3–5] The general method achieves implicit albeit imperfect equalization with a performance ceiling[6–8] that cannot be exceeded without further processing. An example of further processing is to use passive time reversal before explicit equalization.[9,10] In the present study, there are two principle advantages using time-reversal signal processing. First, two key equalizer parameters—the filter length and the update interval—are easily interpreted in terms of the scattering function. Second, because time-reversal processing is inherently based on propagation physics, the results can be used to deduce which channel physics are important for a particular scenario. It is expected that the results for this particular processor will be germane to other forms of equalization.

Equalizer performance results are presented in Sec. III. Data collected on an eight-element receiving array at two different ranges are analyzed. At range of 1 km, optimal

communications performance is achieved when six acoustic paths are retained. By calculating the Rayleigh parameter for rough surface scattering, it is shown that the retained surface-reflected paths have undergone incoherent scattering yet are still of use for coherent communications. Signal loss due to near-surface bubbles appears negligible. Paths with more than one bounce off the sea surface, however, should be rejected. At range of 2 km, ray refraction becomes of increasing importance. Paths that reflect off the sea surface do so at a shallower angle than for the 1 km case, and the Rayleigh parameter is reduced. Consequently, optimal communications performance is achieved when some paths with two sea-surface bounces are retained. The 1 and 2 km results are compared in Sec. IV. Implications for equalizer design are also discussed.

II. COMMUNICATIONS ALGORITHM

In this section, the passive phase conjugation method for coherent underwater communications is sketched. For a more complete, physics-based description of the method, see Refs. 2, 3, and 11. For a signal processing based description, see Refs. 8 and 12. The reader interested only in the algorithm's key design parameters that will be studied in Sec. III using experimental data is referred to this section's last paragraph. In this brief summary, practical implementation issues such as phase tracking and Doppler correction are neglected.

Following Flynn *et al.*,[13] it will be convenient to assume that the signals observed at the receiver have already been brought to baseband and sampled at the symbol rate. Let $y(t)$ represent the signal measured at one element in the Q-element receiving array. The index t represents integer sample time with the sampling taken at the symbol rate. For simplicity, consider binary phase-shift keying (BPSK) so that each symbol $I_l \in \{-1,+1\}$. Then, in general, the measured signal

$$y(t) = \sum_{l=0}^{\infty} h(t,l) I_{t-l} + v_{\text{amb}}(t) \tag{1}$$

is time varying through both the ambient noise ν_{amb} and the complex channel response $h(t,l)$. The goal is to extract the transmitted sequence of symbols. To simplify the processing, it is useful to approximate $h(t,l)$ by the finite-duration, piecewise time-invariant channel response $h_\tau(l)$. The time subscript τ emphasizes that the approximation will apply only for some finite duration, taken here to be N symbols long. As time advances and the acoustic channel changes, it will prove necessary to update $h_\tau(l)$. In this formulation, Eq. (1) is replaced with

$$y(t) = \sum_{l=0}^{L-1} h_\tau(l) I_{t-l} + v(t,\tau). \tag{2}$$

In addition to the ambient noise, $\nu(t,\tau)$ includes two other sources of error: the *truncation error* from limiting the channel response to be L symbols in duration, and the *modeling error* from treating the channel response as being time invariant over a period of N symbols in duration.

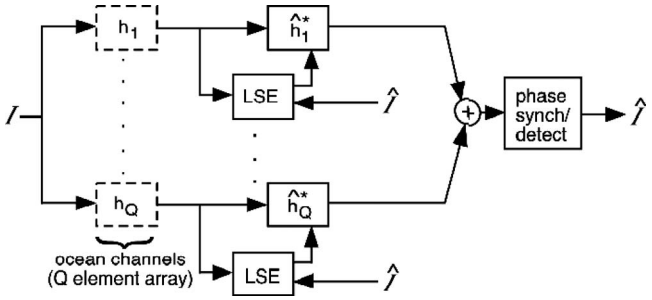


FIG. 1. Baseband-equivalent of decision-directed passive phase conjugation processing for acoustic communication. Data symbols I transmitted through ocean and received on Q-element receiving array. Estimated quantities designated using carets. Least-squares estimation used to produce matched filter $\hat{h}_q$ approximating true channel response h_q for each element in array.

Passive phase conjugation processing is essentially a two-step process. The first step is to match filter the received signal

$$z(t) = \sum_{l=0}^{L-1} h_\tau^*(l) y(t+l). \tag{3}$$

Combining Eqs. (2) and (3), the matched filter output $z(t)$ can be written in terms of the transmitted symbols

$$z(t) = \sum_{l=-L+1}^{L-1} R_l I_{t-l} + w(t), \tag{4}$$

where R_l is the autocorrelation of the channel response and $w(t)$ is the filtered noise. The channel response is complex, but its autocorrelation will be real at zero lag, $l=0$. The second step is to combine the matched filter outputs across each of the Q elements in the receiving array. Introducing the superscript q as the sensor index in Eq. (4), the summation across all sensors is

$$S(t) = \sum_{q=1}^{Q} z^{(q)}(t). \tag{5}$$

Combining Eqs. (4) and (5) yields

$$S(t) = \sum_{l=-L+1}^{L-1} \bar{R}_l I_{t-l} + \bar{w}(t), \tag{6}$$

where $\bar{R}_l$ and $\bar{w}_l$ are the spatially summed autocorrelation and filtered noise, respectively. Because the acoustic multipath structure is different at each of the Q sensors, the complex $\bar{R}_l$ will be real and sharply peaked at $l=0$.[14] In this way Eq. (6) achieves implicit, albeit imperfect equalization.[6–8] The combined signal $S(t)$ is the soft demodulation output that a quantizer then uses to make hard decisions as to the sequence of symbols I_t that were transmitted.

Equations (3)–(5) describe the passive phase conjugation algorithm. The remaining task is to estimate the finite impulse response matched filter $h_\tau^*(l)$ used in Eq. (3), which can be done in a number of ways. One approach is simply to preface the data stream with an isolated probe pulse whose measured response is then used as the matched filter.[3] A more robust approach is to use past decisions about the symbols to update the matched filter. This decision-directed version of

passive phase conjugation[6,12,13] is shown schematically in Fig. 1. The mathematics, emphasizing the algorithm's key design parameters, is outlined as follows.

It is convenient to rewrite Eq. (2) in vector form. Let $\mathbf{h}_\tau^{(q)}$ represent a column vector of length L containing the sampled channel response for receiving array element q. Similarly, $\mathbf{y}^{(q)}$ is a column vector containing N consecutive samples of the received signal. Then

$$\mathbf{y}^{(q)} = \mathbf{X}_t \mathbf{h}_\tau^{(q)} + \mathbf{v}^{(q)}, \tag{7}$$

where $\mathbf{X}_t$ is the $N \times L$ Toeplitz matrix of past symbols

$$\mathbf{X}_t = \begin{pmatrix} I_{t-N+1} & \cdots & I_{t-N-L+2} \\ \vdots & \ddots & \vdots \\ I_t & \cdots & I_{t-L+1} \end{pmatrix}, \tag{8}$$

and $\mathbf{v}^{(q)}$ is the sampled error. The superscripts in Eq. (7) make explicit the terms that depend on the array element index q; note that the same $\mathbf{X}_t$ applies for all elements. Assume that the first $N+L-1$ consecutive symbols, sufficient to construct $\mathbf{X}_t$, represent a training sequence known at the receiver. From knowledge of $\mathbf{X}_t$ and measurement of $\mathbf{y}^{(q)}$, one can formulate the minimization problem

$$\hat{\mathbf{h}}_\tau^{(q)} = \min \|\mathbf{y}^{(q)} - \mathbf{X}_t \mathbf{h}_\tau^{(q)}\|^2. \tag{9}$$

Equation (9) can be solved in a computationally efficient manner.[13] The resulting estimate for the matched filter can be used in Eqs. (3)–(6) to recover the symbols sent after the training sequence.

As the acoustic channel changes in time, the initial estimate for $\hat{\mathbf{h}}_\tau^{(q)}$ will eventually no longer correlate well with the received signal. Suppose $\hat{\mathbf{h}}_\tau^{(q)}$ is used to demodulate M new symbols transmitted after the initial training sequence. These most recent symbols can be used to construct a new $\mathbf{X}_t$ and solve a new minimization problem (9). The resulting new $\hat{\mathbf{h}}_\tau^{(q)}$ is used to recover M additional symbols, and so on through the complete data transmission. After the initial training sequence where the symbols are known, note that the subsequent $\mathbf{X}_t$ must be constructed using the estimated symbols output from the quantizer. Flynn *et al.*[13] showed that the algorithm was robust to errors in $\mathbf{X}_t$. It was also shown that a reasonable value for N, the number of samples used in minimization equation (9), is between $2L$ and $3L$.

The above algorithm has three key design parameters: the number of symbols L defining the length of the matched filter, the number of symbols M that are estimated before refreshing the matched filter, and the number of symbols N over which the channel is assumed to be time invariant and least-squares problem (9) is solved. If L is too small, usable signal is being neglected with consequent degradation in performance. Performance will also be degraded, however, if L is too large as it means what is effectively noise is being included in the matched filter. Intuitively, the refresh interval M is related to rate at which the acoustic arrival structure changes due to changes in the environment. In terms of the scattering function, L is related to the time spread while M is related to the frequency spread. The third parameter N should be selected consistent with the choices for L and M. In Sec. III, the effect of varying these parameters is explored using data collected off the coast of Kauai. It is shown how the optimal values can be related to the underlying physics governing the acoustic propagation and to the local environmental conditions.

III. ALGORITHM PERFORMANCE

The 2003 Kauai experiment (KauaiEx) was designed to study acoustic propagation in the 8 to 50 kHz band. The experiment was conducted in June–July 2003 near the Hawaiian Island of Kauai. Multiple assets were distributed along a 6 km track on a 100 m isobath including several acoustic arrays and numerous environmental sensors. The acoustic transmissions included a variety of communications signals with various modulation schemes. For a general overview of the experiment, see Ref. 14.

For testing the communications algorithm of Sec. II, data in the 8–16 kHz band were used. Using the entire bandwidth, BPSK communications sequences were transmitted at symbol rate of 2174 Hz. Also of interest were linear frequency-modulated (LFM) chirps in the same band. The chirps were useful for estimating the channel impulse response and for interpreting the communications results. The 50 ms duration LFM chirps were repeated every 250 ms for 20 s. Both types of signals were transmitted from a bottom-mounted unit operating at 183 dB re 1 μPa at 1 m.[15] The transmitter was moved to different positions along the isobath over the course of the experiment.

The present analysis is confined to acoustic data collected on a single array deployed midwater column with relatively dense sampling. The eight-element vertical array was moored at 22 08.7734N 159 48.0421W for the duration of the experiment. The array's ITC-6050C transducers (International Transducer Corporation) were uniformly spaced 4.0 m apart with the top element at nominal depth of 22 m. Data were sampled at 50 kHz and stored in a processing unit at depth of 12.5 m. The surface expression of the array included a changeable battery pack and a radio-frequency (rf) modem. Data snippets were sent to the R/V REVELLE via the rf modem to assess data quality and adjust array gains.

An 18-h data set was collected with the range between the transmitter and the array being 1 km and a 16-h data set was collected at range of 2 km. The relevant propagation physics changes with range with consequent effect on communications performance.

A. Performance at range of 1 km

Figure 2 shows environmental data collected beginning 21:00 UTC July 1, 2003. The 18-h period shown coincides with the period while the acoustic array was at range of 1 km. The top panel shows the wind speed and direction. Typical of Hawaii, the wind speed increases in the late morning (local time) and decreases in the evening. Note, however, that there are no completely quiescent periods as the wind speed is always at least 7 m/s. The surface wave spectra in the middle panel show three regimes. Frequencies less than 0.1 Hz correspond to open ocean swell. Frequencies from 0.1 to 0.2 Hz reflect large scale, wind-driven waves consistent

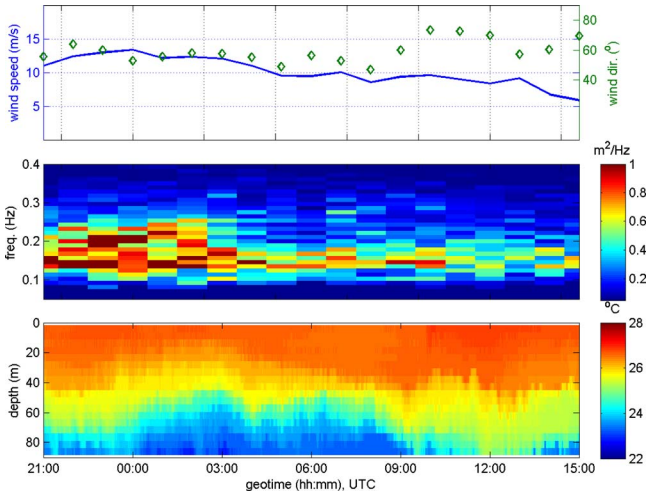


FIG. 2. (Color online) KauaiEx environmental data collected during 18-h period beginning 21:00 UTC July 1, 2003 while receiving array at range of 1 km. Top panel: wind speed (solid line) and direction (diamonds). Middle panel: surface wave spectra. Bottom panel: temperature profile.

with a fully developed sea. Frequencies over 0.2 Hz correspond to surface chop and disappear when the local wind speed drops. By integrating the spectra, the variance of the surface wave height as a function of time is derived. The corresponding significant wave height varies between 1.6 and 1.0 m for the period shown. The bottom panel shows the temperature profile. The surface water is warm and well mixed typically to about 50 m depth. The surface layer is thinned, though, by deep cold water that arrives with the tide; see the period beginning around 0:00. For a more complete description of the environmental data collected over the duration of the experiment, see Refs. 16 and 17.

To better understand the acoustic data collected during this 18-h period, Fig. 3 shows a ray trace. The calculation uses a representative sound speed profile consistent with the temperature data in Fig. 2. Eigenrays are traced to top element (depth 22 m), fifth element (38 m), and bottom element (50 m) of the receiving array. The calculation places the source 7.5 m above the seabed. Because the source is near the seabed, sound gets reflected off the bottom and the eigenrays arrive in pairs. The first arrival pair includes the direct (D) and bottom-bounce (B) paths, the second pair includes the sea-surface (S) and the bottom-then-surface (BS) paths, and the third the SB and BSB paths. Beginning with the fourth pair of arrivals, the acoustic paths have had multiple sea-surface interactions. At range of 1 km, refraction is not a major factor as ray curvature is plainly evident only in the direct path and the path with a single bottom bounce.

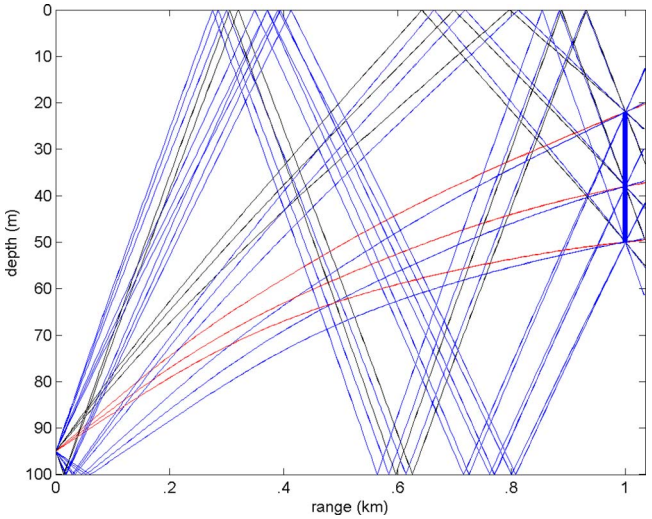


FIG. 3. (Color online) Ray trace to eight-element receiving array at range of 1 km. Eigenrays calculated to top element (depth 22 m), fifth element (38 m), and bottom element (50 m).

A useful metric for characterizing the sea-surface-reflected paths in Fig. 3 is the Rayleigh parameter for rough surface scattering[18]

$$P = 2k\sigma \sin\theta, \tag{10}$$

where k is the acoustic wavenumber, σ is the standard deviation of the surface roughness, and θ is the grazing angle of the incident ray. Each term in Eq. (10) can be calculated using environmental data collected during the experiment: k from the center frequency and the sound speed at the sea surface, σ from the surface wave spectrum (Fig. 2), and θ from the ray trace. For time 21:00 UTC, the Rayleigh parameters for the second, third, and fourth arrival pairs at the middle of the receiving array are 4.4, 8.3, and 10.6, respectively. As observed by Eckart,[19] a Rayleigh parameter of 2 is sufficient to cause a loss in coherent reflection of more than 15 dB. The energy that is lost from the coherently reflected beam is redistributed into the incoherently scattered field. In the context of communications, Kilfoyle and Baggeroer[1] commented that a large Rayleigh parameter implies differential range spreading and incoherent multipath interference. Of present interest is the extent to which these incoherently scattered surface-bounce paths can be used to do coherent underwater communications.

The persistently high wind speeds produce another potential complication in addition to incoherent scattering from the sea surface. The wind speeds exceed the Beaufort velocity[20] for the production of whitecaps with the resulting injection of bubbles into the water column. Assemblages of bubbles can act as a loss mechanism that absorbs energy and reduces the strength of the incoherently scattered paths.

Figure 4 shows the channel response derived from chirp data transmitted at 21:00 UTC. The responses from consecutive match filtered chirps are stacked to show how the detailed arrival pattern varies over 18 s at a single element in the receiving array. The basic arrival pattern is consistent with the ray trace. The first arrival pair with the D and B paths is strong and distinct over the entire period. The next arrival pair with the S and BS paths exhibits the time spread typical of incoherent scattering.[21] Subsequent arrivals show still greater time spread.

The mean-squared error (MSE) in the soft demodulation output is a standard metric for assessing communications performance.[22] The MSE measures the error between the discrete transmitted symbols I and continuous soft demodulation output that is used to estimate the discrete $\hat{I}$; see Fig. 1. Figure 5 shows the MSE for a BPSK sequence transmitted 100 s after the chirp data shown in Fig. 4. The plot shows how the design parameters in the communications algorithm

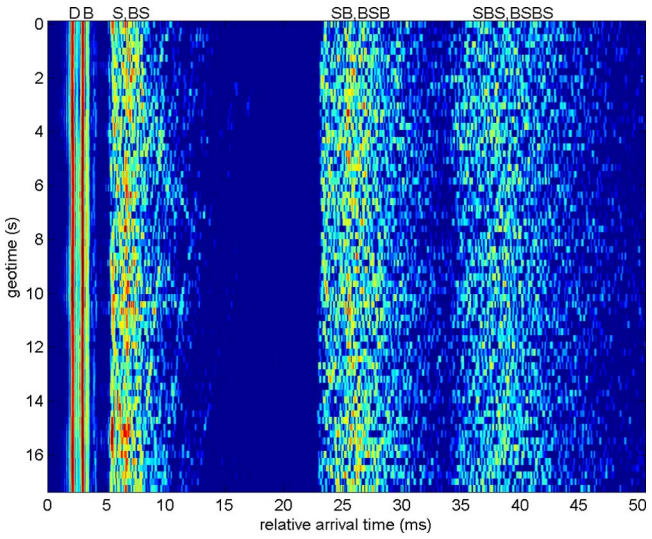


FIG. 4. (Color online) Time-varying channel response. Range from transmitter is 1 km and depth of receiver is 26 m. Result is matched filter output for transmitted LFM signal. Direct (D), bottom-bounce (B), surface-bounce (S), and multiple bounce paths labeled. Scale has 25 dB dynamic range. Data transmitted at 21:00 UTC July 1, 2003.

affect performance. The MSE is plotted versus L, the length of the matched filter expressed in units of symbols. Results are presented for various values of M, the number of symbols extracted before updating the matched filters. The number of samples used to solve channel estimation problem (9) is $N=3L$.

Consider first the results for $M=75$. Performance improves as L is increased up to 20 symbols. At data rate of 2174 symbols/s, $L=20$ symbols is of sufficient duration (9.5 ms) to capture the first four acoustic paths (D, B, S, and BS); see Fig. 4. Increasing the matched filter length from $L=20$ to $L=46$ symbols offers no improvement in communications performance. This corresponds to the period from 9.5 to 21.2 ms in Fig. 4 where there are no arriving paths. Increasing L from 46 to 60 symbols allows two more acoustic ray paths (SB and BSB) to be included in the processing, and the MSE decreases accordingly. Quantitatively, increasing the number of retained paths from four to six has reduced the MSE from -10.2 to -12.2 dB, an improvement of 2.0 dB.

The observed performance improvement by retaining

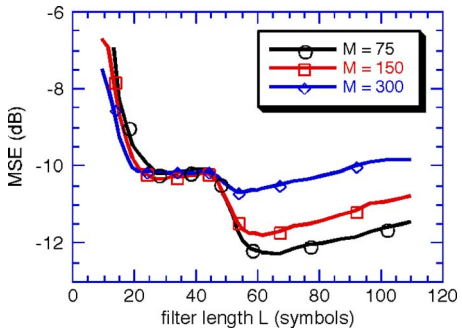


FIG. 5. (Color online) Effect of varying equalizer parameters on communications performance at range of 1 km. MSE of soft demodulation output plotted versus length of matched filter L expressed in units of symbols. Results shown for various update intervals M, also expressed in units of symbols. Data transmitted at 21:01:40 UTC July 1, 2003.

additional paths can be compared to model predictions. The performance model developed by Rouseff[23] predicts

$$10\log[(6\ \text{paths})/(4\ \text{paths})] = 1.8\ \text{dB} \tag{11}$$

improvement in going from four to six paths, in good agreement with the data when $M=75$. The model assumes an isovelocity water column and that each ray path has equal energy. The model further neglects the time spread clearly evident in the channel response data shown in Fig. 4 and ignores any loss due to bubbles. Despite these limitations, the model produces predictions in good agreement with the data for this particular case. Although the paths with a single surface bounce get spread in time, the time-reversal processor is evidently able to recompress this part of the data and achieve the predicted performance improvement; these incoherently scattered paths are of use for coherent communications. Furthermore, there is no evidence of signal loss due to bubbles.

With $M=75$, the matched filters are updated every 34.5 ms. Figure 5 also shows results with longer update intervals. If $L=20$ and only four acoustic paths retained, the performance is essentially unchanged with update interval with $M=300$. Increasing the update interval as high as $M=500$ (not shown) yields little degradation in performance. These first four paths are stable and it is unnecessary to update the matched filters frequently on their behalf. If $L=60$ and six paths are retained, however, $M>75$ yields performance improvement less than the 1.8 dB predicted by Eq. (11). By $M=500$ the performance is actually worse with six paths than it is with four paths. Including these extra two sea-surface interacting paths requires both a longer channel response and more frequent updating. Since the number of symbols N used to solve Eq. (9) is based on the choice for L, the longer channel response also increases the computational burden associated with solving the estimation problem.

Figure 5 further shows that the performance gets worse if the matched filter is lengthened much beyond $L=70$. Ray tracing and the channel response shown in Fig. 4 suggest that values $L>70$ begin to admit the next pair of acoustic paths. Note that these next two paths have been twice reflected off the sea surface. For the present experiment with large Rayleigh parameters at range of 1 km, optimal communications performance is achieved when paths with multiple surface bounces are treated as noise rather than as useful signal.

The results in Fig. 5 are typical at range of 1 km. Figure 6 shows communications performance over the 18-h duration of the experiment using data collected every 0.5 h. The four- and six-path results use $L=30$ and $L=67$, respectively. The results use $N=3L$ symbols to estimate the channel and update the estimate every $M=75$ symbols. Over the 18 h, the MSE varies by less than 3 dB when six paths are retained. For a given number of retained paths, there is no obvious correlation between MSE and the wind speed shown in Fig. 2 although it is again stressed that it was always windy during the experiment. The four- and six-path results track one another with an average of 1.8 dB improvement by including the two extra paths, consistent with Eq. (11).

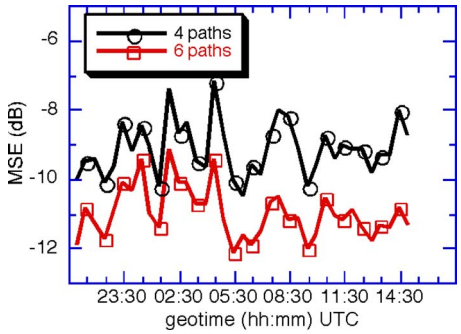


FIG. 6. (Color online) Communications performance at range of 1 km for 18-h period. Effect of varying paths retained by equalizer on MSE in soft demodulation output. Results for same 18-h period shown in Fig. 2. Four- and six-path results retain one reflection off sea surface and use matched filters of lengths $L=30$ and $L=67$, respectively.

B. Performance at range of 2 km

Data were collected with the receiving array at range of 2 km for 16 h starting 07:31 UTC 28 June 2003.

The acoustic arrival structure at range of 2 km is significantly different than what is observed at range of 1 km. Temperature measurements on 28 June show a strong surface mixed layer similar to that shown in Fig. 2. The sound speed contrast between the warm surface water and the cooler water below is sufficient to cause appreciable refraction of the acoustic rays at range of 2 km. As a result, the top of the receiving array may observe different types of reflected and refracted rays than the bottom. Figure 7 shows a ray trace where again eigenrays are traced to top, fifth, and bottom elements of the receiving array. Refraction of the direct and bottom-bounce paths is particularly evident. For this first pair of arrivals, notice that the eigenrays going to the bottom of the receiving array have already passed through a turning point while the eigenrays are still traveling upward at the top of the receiving array. Another factor is that the strongest effects of ocean internal waves and turbulence occur in the vicinity of a turning point at the frequencies and ranges relevant to communications.[24] Furthermore, the details of the arrival pattern will change as the depth of the surface mixed layer rises and falls with the tide.

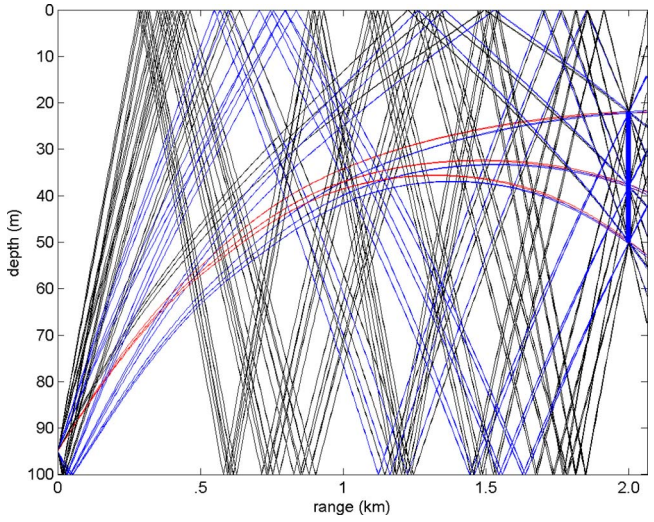


FIG. 7. (Color online) Ray trace to eight-element receiving array at range of 2 km. Eigenrays calculated to top element (depth 22 m), fifth element (38 m), and bottom element (50 m).

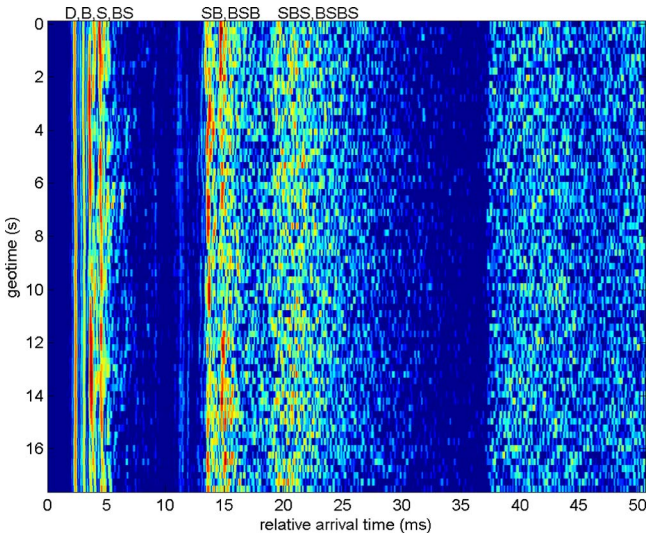


FIG. 8. (Color online) Time-varying channel response. Range from transmitter is 2 km and depth of receiver is 26 m. Result is matched filter output for transmitted LFM signal. Paths labeled as in Fig. 4. Scale has 25 dB dynamic range. Data transmitted at 21:00 UTC June 28, 2003.

The direct and bottom-bounce paths in Fig. 7 suggest sensitivity to the initial launch angle of the rays; small changes in launch angle mean the difference between rays that have or have not passed through a turning point. In such a scenario, the rays would be expected to be relatively low intensity. This was confirmed in Ref. 14 by examining the response to a single chirp at each element in the receiving array. Figure 8 shows this in another way using several consecutive chirps as observed at a single element in the array. Comparing Fig. 8 to Fig. 4, the first arrival pair (D and B) is weaker relative to the second pair (S and BS). At 2 km, there is also less temporal distinction between the first two arrival pairs. This blurring together of arrivals is even more evident between the third (SB and BSB) and fourth (SBS and BSBS) pairs; as the range increases, the arrival structure becomes more compact.

Figure 9 shows a sample result of communications algorithm performance at range of 2 km. As with Fig. 6, the MSE

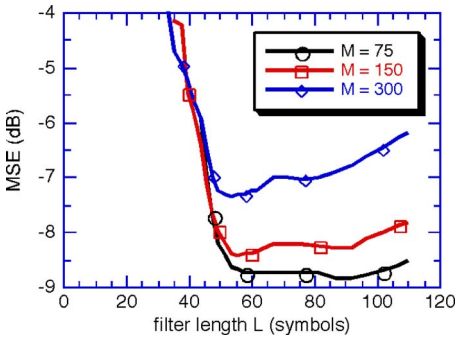


FIG. 9. (Color online) Effect of varying equalizer parameters on communications performance at range of 2 km. MSE of soft demodulation output plotted versus length of matched filter L expressed in units of symbols. Results shown for various update intervals M, also expressed in units of symbols. Data transmitted at 21:01:40 UTC June 28, 2003.

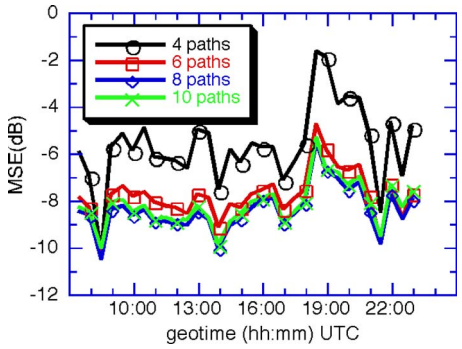


FIG. 10. (Color online) Communications performance at range of 2 km for 16-h period. Effect of varying paths retained by equalizer on MSE in soft demodulation output. Eight- and ten-path results retain two reflections off sea surface and use matched filter of lengths $L=67$ and $L=110$, respectively.

is plotted versus matched filter length L for various update intervals M. Because the arrivals are more blurred together (Fig. 8), the distinct performance levels observed at 1 km are not evident. Because the different arrivals are of different strengths, a simple scaling rule for performance like Eq. (11) does not apply.

With filter length $L=40$ symbols, sufficient to retain six paths at 2 km, the MSE in Fig. 9 is essentially identical for $M=75$, 150, or 300 symbols. This is in contrast to the 1 km results where $M=75$ was required to get the full benefit of six retained paths. At the longer range, a surface-bounce path is incident on the sea surface at a shallower angle and the Rayleigh parameter is reduced. Evidently, these paths are thus less impacted by sea-surface dynamics and the matched filter needs to be updated less often.

With update interval $M=75$ symbols, the performance improves steadily with increasing filter length to $L\approx 67$. The choice $L=67$, sufficient to include eight paths at 2 km, includes paths with two reflections off the sea surface. In contrast to results at 1 km, the processor is able to treat paths with multiple sea-surface reflections as useful signal. The minimum MSE occurs at $L=90$, a filter sufficiently long to retain two additional paths with multiple sea-surface reflections. The marginal improvement in going from eight to ten paths, however, is slight and is only observed at all with frequent updating of the matched filter.

Figure 10 shows results over the 16-h period where data were collected at range of 2 km. The MSE is plotted versus time for various numbers of retained paths. The communications algorithm could perform using as few as four acoustic paths ($L=20$), but for this choice it was necessary to use $N>3L$ symbols to solve channel estimation problem (9). Consequently, all results in Fig. 10 use $N=200$ and $M=75$. Over the 16 h, performance always improves in going from four to six paths. Performance usually improves in going from six to eight paths, and occasionally improves in going to ten paths. The marginal improvement in adding extra paths decreases. Keeping eight or ten paths implies keeping paths that have reflected twice off the sea surface.

IV. DISCUSSION

The communications results in Sec. III are parametrized in terms of the number of acoustic paths retained by the equalizer as useful signal. Because the acoustic source is near the bottom, the rays arrive in pairs. Retaining four or six paths means rejecting all paths with more than one reflection from the sea surface. Retaining eight or ten paths means keeping paths with two sea-surface reflections. Table I summarizes the experimental results. The results are an average over 18 h at 1 km and 16 h at 2 km. Also included are the Rayleigh parameter calculations using $\sigma=0.34$ m for the surface roughness. For each case at both ranges, the Rayleigh parameter is sufficiently large to imply incoherent scattering.

Consider the 1 km results. In going from four to six paths, communications performance improves by 1.8 dB consistent with Eq. (11). This improvement, however, comes with definite costs: The matched filters must be lengthened and the size of channel estimation problem (9) increases proportionally. Furthermore, these additional paths are more sensitive to the sea surface, as quantified by the Rayleigh parameter, and as a result the filters must be updated much more frequently. The communications engineer could make rational decisions as to whether the increased computational burden is worth the 1.8 dB improvement in performance. It is definitely not advisable to retain paths with more than one sea-surface reflection as the performance actually gets worse.

As the range increases to 2 km, the situation becomes more complicated. The arrival structure becomes more compact with less distinction between the arrival pairs. Ray refraction causes the direct and bottom-bounce paths to become weaker relative to subsequent arrivals. Consequently, in going from four to six paths, the performance improves by more than the 1.8 dB predicted by the simple model. Sound also reflects off the sea surface at shallower angles; the Rayleigh parameter for eight paths at 2 km is less than it is for six paths at 1 km even though the former includes paths with two sea-surface reflections. At the longer range, the equalizer is able to treat these paths with multiple surface reflections as useful signal. The choice $L=67$ symbols for the matched filter length, sufficient for six paths at 1 km and eight paths at 2 km, yields satisfactory results at both ranges.

TABLE I. Comparison of 1 and 2 km results.

Paths kept	Maximum sea-surface bounces	Maximum Rayleigh parameter	1 km filter length, L (symbols)	Average MSE (dB)	Maximum Rayleigh parameter	2 km filter length, L (symbols)	Average MSE (dB)
4	1	4.4	30	−9.2	1.8	20	−5.7
6	1	8.3	67	−11.0	3.8	38	−7.8
8	2	10.6	110	−10.3	5.3	67	−8.4
10	2	⋯	⋯	⋯	7.3	110	−8.3

V. CONCLUDING REMARKS

The performance of an equalizer has been quantified in terms of the acoustic paths retained as useful signal using acoustic data collected during the KauaiEx experiment. Augmenting the acoustic data with concurrent environmental data such as the sound speed profile and the sea-surface conditions proved crucial in interpreting the results.

By examining performance at two different ranges, the roles of reflection and refraction on the usefulness of various acoustic paths have been studied. Because the sea surface was rough, there was negligible coherent reflection, and all acoustic paths that interacted with the sea surface were incoherently scattered. The results show, however, that some incoherently scattered paths might still be useful for coherent communications. At shorter range, paths reflect off the surface at steeper grazing angles and there is more time spread between arrivals than is observed at the longer range. Both of these factors hamper acoustic communication. For the KauaiEx data, consequently, only paths with no more than one sea-surface interaction were useful for communications at range of 1 km. At the more distant range, the grazing angles are shallower and there is less time spread between arrivals. As a result, paths with up to two sea-surface reflections are usable.

While a communications engineer will not usually have access to detailed environmental data, he or she might still be able to make reasonable guesses as to the sea state or to the type of sound speed profile that might be expected for a given scenario. These expectations, together with experience from experiments such as KauaiEx, could be used to guide the design of an equalizer.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research. The authors thank the anonymous reviewers for several useful suggestions.

[1]D. B. Kilfoyle and A. B. Baggeroer, "The state of the art of underwater acoustic telemetry," IEEE J. Ocean. Eng. **25**, 4–27 (2000).
[2]D. R. Dowling, "Acoustic pulse-compression using passive phase-conjugate processing," J. Acoust. Soc. Am. **95**, 1450–1458 (1994).
[3]D. Rouseff, D. R. Jackson, W. L. J. Fox, J. A. Ritcey, and D. R. Dowling, "Underwater acoustic communications by passive phase conjugation: Theory and experimental results," IEEE J. Ocean. Eng. **26**, 821–831 (2001).
[4]T. C. Yang, "Temporal resolution of time-reversal and passive phase conjugation processing," IEEE J. Ocean. Eng. **28**, 229–245 (2003).
[5]J. Gomes, A. Silva, and S. Jesus, "Adaptive spatial combining for passive time-reversed communications," J. Acoust. Soc. Am. **124**, 1038–1053 (2008).
[6]J. A. Flynn, W. L. J. Fox, J. A. Ritcey, and D. Rouseff, "Performance of reduced-complexity multi-channel equalizers for underwater acoustic communications," in *Proceedings of the 36th Asilomar Conference*, Vol. **1**, pp. 453–460 (2002).
[7]J. Preisig, "The impact of underwater acoustic channel structure and dynamics on the performance of adaptive coherent equalizers," in *High Frequency Ocean Acoustics*, edited by M. B. Porter, M. Siderius, and W. A. Kuperman (AIP, New York, 2004).
[8]M. Stojanovic, "Retrofocusing techniques for high rate acoustic communications," J. Acoust. Soc. Am. **117**, 1173–1185 (2005).
[9]H. C. Song, W. S. Hodgkiss, W. A. Kuperman, M. Stevenson, and T. Akal, "Improvement of time-reversal communications using adaptive channel equalizers," IEEE J. Ocean. Eng. **31**, 487–496 (2006).
[10]H. C. Song, W. S. Hodgkiss, and S.-M. Kim, "Performance prediction of passive time reversal communications (L)," J. Acoust. Soc. Am. **122**, 2517–2518 (2007).
[11]A. Silva, S. Jesus, J. Gomes, and V. Barroso, "Underwater acoustic communication using a 'virtual' electronic time-reversal mirror approach," in *Proceedings of the Fifth European Conference on Underwater Acoustics (EC 2000)* (2000), pp. 531–536.
[12]J. A. Flynn, J. A. Ritcey, W. L. J. Fox, D. R. Jackson, and D. Rouseff, "Decision-directed passive phase conjugation: Equalisation performance in shallow water," Electron. Lett. **37**, 1551–1553 (2001).
[13]J. A. Flynn, J. A. Ritcey, D. Rouseff, and W. L. J. Fox, "Multichannel equalization by decision-directed passive phase conjugation: Experimental results," IEEE J. Ocean. Eng. **29**, 824–836 (2004).
[14]M. Porter, P. Hursky, M. Siderius, M. Badiey, J. Caruthers, W. Hodgkiss, K. Raghukumar, D. Rouseff, W. Fox, C. de Moustier, B. Calder, B. Kraft, V. McDonald, P. Stein, J. Lewis, and S. Rajan, "The Kauai experiment," in *High Frequency Ocean Acoustics*, edited by M. B. Porter, M. Siderius, and W. A. Kuperman (AIP, New York, 2004), pp. 307–321.
[15]V. K. McDonald, P. Hursky, and KauaiEx Group, "Telesonar testbed instrument provides a flexible platform of acoustic propagation and communications research in the 8–50 kHz band," in *High Frequency Ocean Acoustics*, edited by M. B. Porter, M. Siderius, and W. A. Kuperman (AIP, New York, 2004), pp. 336–349.
[16]M. Badiey, A. Song, D. Rouseff, H. C. Song, W. S. Hodgkiss, and M. B. Porter, "High-frequency acoustic propagation in the presence of ocean variability in KauaiEx," in *Proceedings of the OCEANS 2007—Europe* (IEEE, New York, 2007), pp. 1–4.
[17]A. Song, M. Badiey, H. C. Song, W. S. Hodgkiss, M. B. Porter, and KauaiEx Group, "Impact of ocean variability on coherent underwater acoustic communications during the Kauai experiment (KauaiEx)," J. Acoust. Soc. Am. **123**, 856-865 (2008).
[18]L. M. Brekhovskikh and Y. P. Lysanov, *Fundamentals of Ocean Acoustics*, 2nd ed. (Springer-Verlag, New York, 1991), pp. 21–23 and 182–183.
[19]C. Eckart, "The scattering of sound from the sea surface," J. Acoust. Soc. Am. **25**, 566–570 (1953).
[20]E. C. Monahan and M. Lu, "Acoustically relevant bubble assemblages and their dependence on meteorological parameters," IEEE J. Ocean. Eng. **15**, 340–349 (1990).
[21]P. H. Dahl, J. W. Choi, N. J. Williams, and H. C. Graber, "Field measurements and modeling of attenuation from near-surface bubbles for frequencies 1–20 kHz," J. Acoust. Soc. Am. **124**, EL163–EL169 (2008).
[22]J. G. Proakis, *Digital Communications*, 3rd ed. (McGraw-Hill, New York, 1995), pp. 607–608.
[23]D. Rouseff, "Intersymbol interference in underwater acoustic communications using time-reversal signal processing," J. Acoust. Soc. Am. **117**, 780–788 (2005).
[24]F. S. Henyey, D. Rouseff, J. M. Grochocinski, S. A. Reynolds, K. L. Williams, and T. E. Ewart, "Effect of internal waves and turbulence on a horizontal aperture sonar," IEEE J. Ocean. Eng. **22**, 270–280 (1997).

Forward propagation of time evolving acoustic pressure: Formulation and investigation of the impulse response in time-wavenumber domain

Vincent Grulier, Sébastien Paillasseur, Jean-Hugh Thomas,[a] and Jean-Claude Pascal
Laboratoire d'Acoustique de l'Université du Maine (LAUM UMR-CNRS 6613) and Ecole Nationale Supérieure d'Ingénieurs du Mans, rue Aristote, 72085 Le Mans Cedex 09, France

Jean-Christophe Le Roux
Centre de Transfert de Technologie du Mans, 20 rue Thales de Milet, 72000 Le Mans, France

(Received 11 March 2009; revised 9 August 2009; accepted 10 August 2009)

The aim of this work is to continuously provide the acoustic pressure field radiated from nonstationary sources. From the acquisition in the nearfield of the sources of a planar acoustic field which fluctuates in time, the method gives instantaneous sound field with respect to time by convolving wavenumber spectra with impulse response and then inverse Fourier transforming into space for each time step. The quality of reconstruction depends on the impulse response which is composed of investigated parameters as transition frequency and propagation distance. Sampling frequency also affects errors of the practically discrete impulse response used for calculation. To avoid aliasing, the impulse response is low-pass filtered with Chebyshev or Kaiser–Bessel filter. Another approach to implement the impulse response consists of applying an inverse Fourier transform to the theoretical transfer function for propagation. To estimate the performance of each processing method, a simulation test involving several source monopoles driven by nonstationary signals is executed. Some indicators are proposed to assess the accuracy of the temporal signals predicted in a forward plane. The results show that the use of a Kaiser–Bessel filter numerically implemented or that of the inverse Fourier transform can provide the most accurate instantaneous acoustic signals. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3216916]

PACS number(s): 43.60.Gk, 43.20.Px, 43.60.Sx [EJS] Pages: 2367–2378

I. INTRODUCTION

Knowledge of the instantaneous acoustic field radiated far from sources is of particular interest in several applications. Among them, active control of the noise radiated by structural bodies requires a monitoring signal in the nearfield or in the farfield to provide appropriate signals to the actuators in order to suppress the noise. In many cases, this control only concerns some components of the radiated field, associated with a spatial representation or with components which can be filtered within the wavenumber spectrum obtained from Fourier transform of the spatial pressure distribution. A representative temporal signal is then necessary for the control algorithm.[1,2] Other applications need a description of the instantaneous radiation (for instance, due to unsteady excitation such as impacts or turbulent flows), whose spatial and time features are of great importance for the acoustic perception of a listener. The applications mainly concern the study of musical instruments, how to model them for sound synthesis,[3,4] although more industrial applications are interested in sound quality. In this field, reducing impact noise is particularly difficult and requires a technique to efficiently characterize the phenomena.[5,6]

We are interested in acquiring or simulating a time-dependent acoustic field using a microphone array in the nearfield of the sources so as to find the time-dependent pressure field in a plane further from the sources. Furthermore, solving this problem would be the first step of a more complex task which consists of reconstructing a nonstationary pressure field directly on a source plane from measurements done in the nearfield as in nearfield acoustic holography involving stationary acoustic sources.[7] The resolution of this inverse problem[8] is not considered in this paper, as the focus is on the direct problem which is forward prediction of acoustic pressure field. In the article, this direct problem is called forward radiation problem.

For harmonic signals (using the Helmholtz integral), a widely used method for calculating the acoustic radiation consists of formulating and solving the problem in the wavenumber domain[7,9] and then transforming back to the spatial domain. To ensure the transition to the spatial domain, some numerical solutions using fast Fourier transforms (FFTs) have been proposed.[10] Among the other possible solutions, working in the wavenumber domain is without doubt the most efficient using the FFT algorithm.[11] With this method, the spatial Fourier transform with respect to both variables x and y of the Helmholtz equation leads to the solution of the equation in the z direction and the prediction of how the wavenumber spectrum propagates in this direction for each angular frequency ω. However, the solution which consists of solving the problem for each spectral component and then synthesizing the temporal signal through an inverse Fourier transform[11] is time consuming and sensitive to errors for the

[a]Author to whom correspondence should be addressed. Electronic mail: jean-hugh.thomas@univ-lemans.fr

low frequency components. Indeed the method requires one Fourier transform with respect to time for each measurement point of the array, one spatial Fourier transform and processing of the wavenumber spectrum for each spectral line, and at last one inverse Fourier transform with respect to time for each measurement point. A fundamentally different approach is presented here which involves the time-wavenumber domain without operating in the frequency domain. After applying a Fourier transform with respect to the planes x and y, an equation along the z axis describes how the instantaneous wavenumber spectrum propagates. This equation is then solved by using the Laplace transform. A similar approach but different in the way of seeking the solution was proposed by Forbes *et al.*[12] providing the same result. The solution can be presented as a convolution product between each component of the wavenumber spectrum and an impulse response obtained analytically in the time-wavenumber domain. This has the advantage of continuously processing the signal so that each new sample picked up by the microphones provides a new sample of the propagated pressure field. This method is the central part of the study presented here where the relevance of the impulse response is tested from nonstationary simulated acoustic sources. The main question to answer is how to implement the impulse response and with what sample rate. We also emphasize the fact that the impulse response should be processed before projecting the input pressure field. Some criteria are finally given to assess the viability of the method and to compare different processing methods applied to the impulse response in the time-wavenumber domain.

The theoretical formulation of the method providing an impulse response in the time-wavenumber domain is given in Sec. II. In spite of the fact that the starting equation is also the wave equation, the presentation of the approach based here on the Laplace formalism differs from that of Forbes *et al.*[12] but leads to the same expression. Then the shape and the frequency response of the impulse response are presented in Sec. III. In particular, an important feature of the impulse response is highlighted. It is the transition frequency which separates for each point of the wavenumber domain two kinds of travelling waves, propagating or decaying. Next, the aim is to implement the impulse response using a finite number of samples. The influence of several parameters such as the propagation distance, the transition frequency, and the sampling frequency is investigated in Sec. IV. Then several processing methods to implement an effective impulse response are described. Most of them are based on low-pass filtering of the response. An approach using the inverse Fourier transform is also mentioned. Numerical results are discussed in Sec. V while the source plane is composed of three monopoles generating nonstationary acoustic signals. Some indicators[13,14] are given to objectively compare the signals forward propagated to another plane.

II. WAVE EQUATION SOLUTION IN TIME-WAVENUMBER DOMAIN

First, the wave equation in Cartesian geometry is considered:

$$\nabla^2 p(x,y,z,t) - \frac{1}{c^2}\frac{\partial^2 p(x,y,z,t)}{\partial t^2} = 0. \tag{1}$$

By applying the Fourier transform with respect to x and y to Eq. (1), using the time-wavenumber spectrum $P(k_x,k_y,z,t)$ given by

$$P(k_x,k_y,z,t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y,z,t)e^{j(k_x x+k_y y)}dxdy, \tag{2}$$

Eq. (1) yields

$$\frac{\partial^2 P(k_x,k_y,z,t)}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 P(k_x,k_y,z,t)}{\partial t^2} - (k_x^2+k_y^2)P(k_x,k_y,z,t) = 0. \tag{3}$$

By setting

$$F(z,t) = P(k_x,k_y,z,t), \quad a = \frac{1}{c^2}, \quad b = k_x^2 + k_y^2, \tag{4}$$

Eq. (3) can be rewritten as

$$\frac{\partial^2 F(z,t)}{\partial z^2} - a\frac{\partial^2 F(z,t)}{\partial t^2} - bF(z,t) = 0. \tag{5}$$

To seek the solution of Eq. (3) for each point (k_x,k_y) of the instantaneous wavenumber spectrum, the Laplace formalism[15] is used with $\mathcal{L}(F(z,t))=f(z,s)$ the Laplace transform of $F(z,t)$. Considering

$$\mathcal{L}\left(\frac{\partial^2 F(z,t)}{\partial z^2}\right) = \frac{\partial^2 f(z,s)}{\partial z^2},$$

$$\mathcal{L}\left(\frac{\partial F(z,t)}{\partial t}\right) = sf(z,s) - F(z,0),$$

$$\mathcal{L}\left(\frac{\partial^2 F(z,t)}{\partial t^2}\right) = s^2 f(z,s) - sF(z,0) - \frac{\partial F(z,0)}{\partial t}, \tag{6}$$

and that the initial condition $(t=0)$ is zero, as the solution sought is an impulse response,

$$F(z,0) = 0, \tag{7}$$

the Laplace transform of Eq. (5) yields

$$\frac{\partial^2 f(z,s)}{\partial z^2} - (as^2+b)f(z,s) = 0. \tag{8}$$

The general solution of Eq. (8) is

$$f(z,s) = Ke^{\alpha z}, \tag{9}$$

where K and α are two constants which need to be found.

Substituting Eq. (9) into Eq. (8) leads to two possible values for α:

$$\alpha = \pm\sqrt{as^2+b}. \tag{10}$$

Since waves propagate toward the increasing z axis and since it is assumed that there is no reflective wave, the solution chosen for α is the negative one,

$$\alpha = -\sqrt{as^2+b}. \tag{11}$$

K is given by the initial condition for $z=0$,

$$K = f(0,s). \tag{12}$$

Hence after substituting Eqs. (11) and (12) into Eq. (9), the solution for $f(z,s)$ becomes

$$f(z,s) = f(0,s)e^{-z\sqrt{as^2+b}}. \tag{13}$$

According to Hladik,[15] the exponential term $e^{-z\sqrt{as^2+b}}$ in Eq. (13) can be expressed as

$$e^{-z\sqrt{as^2+b}} = u(z,s) + v(z,s), \tag{14}$$

where

$$u(z,s) = e^{z\sqrt{as^2}} \tag{15}$$

and

$$v(z,s) = -z\sqrt{b}\int_{z\sqrt{a}}^{\infty} e^{-st}\frac{J_1(\sqrt{b/a}\sqrt{t^2-az^2})}{\sqrt{t^2-az^2}}dt, \tag{16}$$

where J_1 denotes the Bessel function of the first kind and order 1.

Hence Eq. (13) yields

$$f(z,s) = f(0,s)u(z,s) + f(0,s)v(z,s). \tag{17}$$

Considering

$$\mathcal{L}(F(0,t)) = f(0,s),$$

$$\mathcal{L}(V(z,t)) = v(z,s),$$

$$\mathcal{L}(F(z,t-t_0)) = f(z,s)e^{-st_0}, \tag{18}$$

and applying the inverse Laplace transform to Eq. (17), it follows that

$$F(z,t) = \begin{cases} 0 & \text{for } 0 \le t < z\sqrt{a} \\ F(0,t-z\sqrt{a}) + F(0,t) * V(z,t) & \text{for } t \ge z\sqrt{a}, \end{cases} \tag{19}$$

with

$$V(z,t) = \begin{cases} 0 & \text{for } 0 \le t < z\sqrt{a} \\ -z\sqrt{b}\dfrac{J_1(\sqrt{b/a}\sqrt{t^2-az^2})}{\sqrt{t^2-az^2}} & \text{for } t \ge z\sqrt{a}. \end{cases} \tag{20}$$

Since we used the notation given by Eq. (4), hence with $\sqrt{a}=1/c$ and $\sqrt{b}=\sqrt{k_x^2+k_y^2}$, the solution of Eq. (3) for the time-wavenumber spectrum is provided from Eqs. (19) and (20):

$$P(k_x,k_y,z,t) = 0 \quad \text{for } 0 \le t < \frac{z}{c}$$

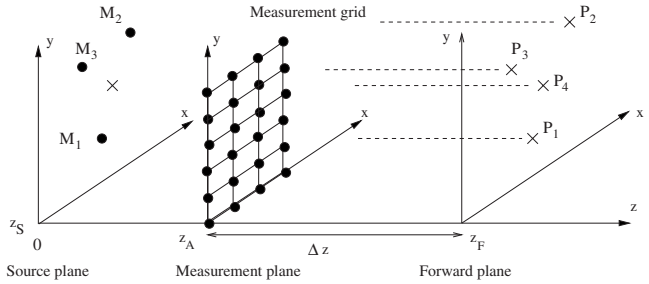


FIG. 1. Geometry of the radiation problem: the pressure field in $z=z_F$ has to be computed from the pressure field acquired in $z=z_A$. The numerical study involves three nonstationary monopoles M_1, M_2, and M_3.

$$P(k_x,k_y,z,t) = P\left(k_x,k_y,0,t-\frac{z}{c}\right) - P(k_x,k_y,0,t) * \left[z\sqrt{k_x^2+k_y^2}\frac{J_1\left(c\sqrt{k_x^2+k_y^2}\sqrt{t^2-\frac{z^2}{c^2}}\right)}{\sqrt{t^2-\frac{z^2}{c^2}}}\Gamma\left(t-\frac{z}{c}\right)\right]$$

$$\text{for } t \ge \frac{z}{c}. \tag{21}$$

Γ is the Heaviside function defined by

$$\Gamma(t) = \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{2} & \text{for } t = 0 \\ 1 & \text{for } t > 0. \end{cases} \tag{22}$$

For $t \ge z/c$, Eq. (21) can be expressed as a convolution product between the time-wavenumber spectrum $P(k_x,k_y,0,t)$ and an impulse response $h(k_x,k_y,z,t)$:

$$P(k_x,k_y,z,t) = P(k_x,k_y,0,t) * h(k_x,k_y,z,t), \tag{23}$$

where $h(k_x,k_y,z,t)$ is given by

$$h(k_x,k_y,z,t) = \delta\left(t-\frac{z}{c}\right) - z\sqrt{k_x^2+k_y^2}\frac{J_1\left(c\sqrt{k_x^2+k_y^2}\sqrt{t^2-\frac{z^2}{c^2}}\right)}{\sqrt{t^2-\frac{z^2}{c^2}}}\Gamma\left(t-\frac{z}{c}\right). \tag{24}$$

$\delta(t)$ denotes the Dirac distribution.

III. FORWARD PROPAGATION OF TIME EVOLVING PRESSURE FIELD

A. Propagation in the time-wavenumber domain

By considering the geometry of the problem (see Fig. 1), the time-dependent wavenumber spectrum $P(k_x,k_y,z_F,t)$ in a forward plane $z=z_F$ can be obtained by convolving each component of the time-dependent wavenumber spectrum $P(k_x,k_y,z_A,t)$ acquired in a measurement plane $z=z_A$ with an impulse response $h(k_x,k_y,z_F-z_A,t)$ in the time-wavenumber domain:

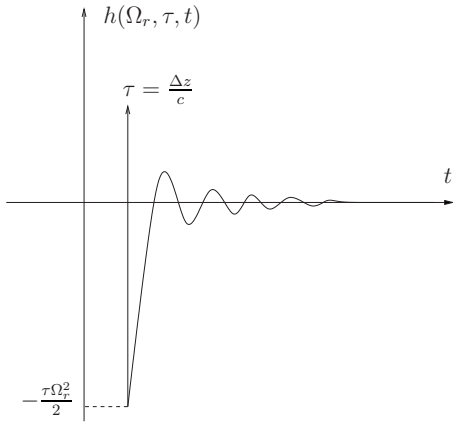


FIG. 2. Shape of the impulse response $h(\Omega_r,\tau,t)$ defined in Eq. (26).

$$P(k_x,k_y,z_F,t)=P(k_x,k_y,z_A,t)*h(k_x,k_y,z_F-z_A,t). \quad (25)$$

By using the following notation for the propagation distance $\Delta z=z_F-z_A$, the wavenumber $k_r=\sqrt{k_x^2+k_y^2}$, the propagation delay $\tau=\Delta z/c$, and the transition pulsation $\Omega_r=ck_r$, the impulse response $h(k_x,k_y,\Delta z,t)$ of Eq. (24) can be rewritten as

$$h(\Omega_r,\tau,t)=\delta(t-\tau)-\tau\Omega_r^2\frac{J_1(\Omega_r\sqrt{t^2-\tau^2})}{\Omega_r\sqrt{t^2-\tau^2}}\Gamma(t-\tau). \quad (26)$$

The theoretical impulse response $h(\Omega_r,\tau,t)$ is represented in Fig. 2. In practical applications, the instantaneous wavenumber spectrum is obtained by processing a two-dimensional spatial Fourier transform at each discrete time of the signals acquired by a microphone array. The pressure field $p(x,y,z,t)$ on the plane $z=z_F$ is obtained by processing an inverse two-dimensional spatial Fourier transform of the result of Eq. (25). The discrete problem is studied in Sec. IV.

B. Interpretation in the frequency domain

The frequency response $H(\Omega_r,\tau,f)$ is the Fourier transform with respect to time of the impulse response $h(\Omega_r,\tau,t)$. It can also be highlighted by applying a Fourier transform to Eq. (25). The obtained equation is then

$$P(k_x,k_y,z_F,\omega)=P(k_x,k_y,z_A,\omega)H(\Omega_r,\tau,\omega). \quad (27)$$

This describes the relationship between the known pressure field on a plane $z=z_A$ and the pressure on any other plane $z=z_F$ when the studied stationary acoustic sources are confined on the half plane $z\leq z_S$ ($z=z_S$ is the source plane),[7]

$$P(k_x,k_y,z_F,\omega)=P(k_x,k_y,z_A,\omega)G_P(k_r,\Delta z,\omega), \quad (28)$$

where the propagator G_P is defined by

$$G_P(k_r,\Delta z,\omega)=e^{-jk_z\Delta z}=\begin{cases}e^{-j\Delta z\sqrt{(\omega/c)^2-k_r^2}} & \text{for } \dfrac{\omega}{c}\geq k_r\\ e^{-\Delta z\sqrt{k_r^2-(\omega/c)^2}} & \text{for } \dfrac{\omega}{c}<k_r,\end{cases} \quad (29)$$

where c is the sound speed.

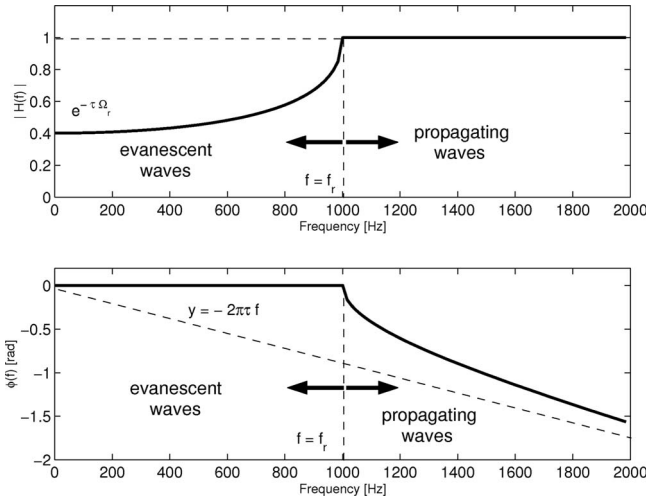


FIG. 3. Magnitude and phase in rad of the theoretical transfer function $H(\Omega_r,\tau,f)$ for the transition frequency $f_r=1000$ Hz and the propagation distance $\Delta z=0.05$ m.

By using Eqs. (27)–(29), the frequency response $H(\Omega_r,\tau,\omega)$ can be written as

$$H(\Omega_r,\tau,\omega)=G_P(\Omega_r,\tau,\omega)=\begin{cases}e^{-j\tau\sqrt{\omega^2-\Omega_r^2}} & \text{for } \omega\geq\Omega_r\\ e^{-\tau\sqrt{\Omega_r^2-\omega^2}} & \text{for } \omega<\Omega_r.\end{cases} \quad (30)$$

The magnitude and the phase of the frequency response are represented in Fig. 3 for a transition frequency $f_r=\Omega_r/2\pi=1000$ Hz and $\Delta z=0.05$ m. Examining this figure, it is clear that f_r is a transition frequency. It is the frequency that separates two kinds of behavior for the acoustic waves: propagating waves for $f\geq f_r$ and exponentially decaying sound fields (evanescent waves) for $f<f_r$. The nonstationary signal in the time-wavenumber domain $P(k_x,k_y,z_A,t)$, which is the time evolving pressure in the plane $z=z_A$ at the point k_r of the wavenumber spectrum, will show that its frequency components above the transition frequency propagate as propagating waves and its frequency components below f_r decay exponentially.

IV. DISCRETE FINITE LENGTH IMPULSE RESPONSE

The forward radiation problem is solved by using a convolution in the time-wavenumber domain [see Eq. (25)] between the input acoustic signals and the impulse response in Eq. (26). The accuracy of the instantaneous radiating sources reconstructed from the measurements depends on the sampling rate of the analytical impulse response. It is then very important that the discrete Fourier transform $H_F(\Omega_r,\tau,\omega)$ of the sampled impulse response is close to the theoretical transfer function $H(\Omega_r,\tau,\omega)$. Since the sampled impulse response is not band limited, one more processing stage involving a low-pass filter is added to avoid aliasing effect.

A. Frequency analysis of the sampled impulse response

Here the influence on the sampled impulse response of both parameters Δz the distance between the measurement

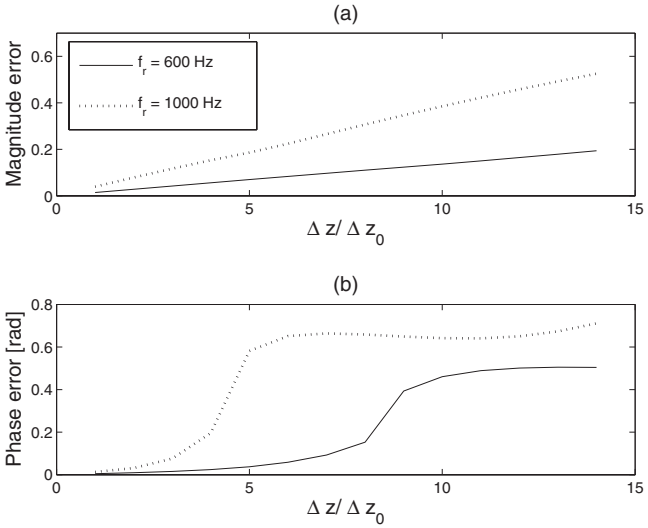


FIG. 4. Errors (a) in the magnitude $E_{|H_F|}$ and (b) in the phase E_{ϕ_F} (rad) of the transfer function $H_F(f)$ as functions of the normalized propagation distance $\Delta z/\Delta z_0$ with $\Delta z_0 = c\Delta t = 0.0215$ m. The transfer function is obtained by applying a Fourier transform to the sampled impulse response in Eq. (26) for $f_e = 16\,000$ Hz and 256 samples.

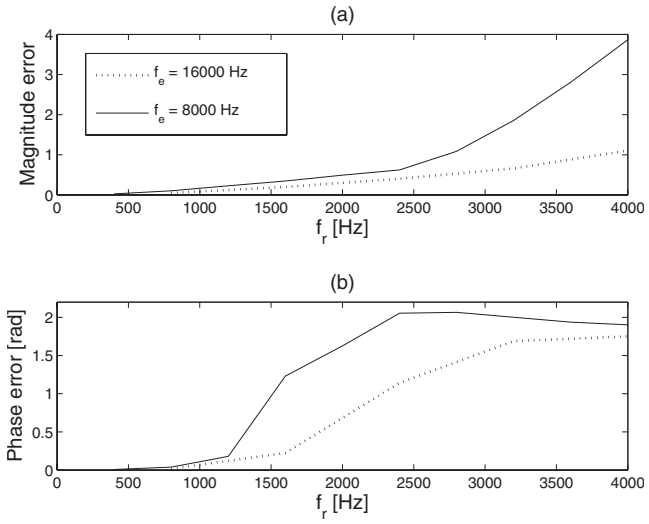


FIG. 5. Errors (a) in the magnitude $E_{|H_F|}$ and (b) in the phase E_{ϕ_F} (rad) of the transfer function $H_F(f)$ as functions of the transition frequency ($\Delta z = 0.043$ m) for two sampling frequencies $f_e = 8000$ Hz ($\Delta z/c\Delta t = 1$) and $f_e = 16\,000$ Hz ($\Delta z/c\Delta t = 2$). 256 samples are considered.

plane and the forward plane, and f_r the transition frequency is investigated. The sampling frequency is f_e and the time interval between successive samples is Δt, the sampling period with $\Delta t = 1/f_e$. The discrete Fourier transform $H_F(\Omega_r, \tau, \omega)$ of the 256 samples of the impulse response $h(\Omega_r, \tau, t)$ is then compared to $H(\Omega_r, \tau, \omega)$. For this purpose, two root mean square errors $E_{|H_F|}$ and E_{ϕ_F} for both the magnitude and the phase of $H_F(\Omega_r, \tau, f) = |H_F(f)|e^{j\phi_F(f)}$ are calculated. They are defined by

$$E_{|H_F|} = \sqrt{\langle (|H(f)| - |H_F(f)|)^2 \rangle}, \tag{31}$$

$$E_{\phi_F} = \sqrt{\langle (\phi(f) - \phi_F(f))^2 \rangle}, \tag{32}$$

where $|H(f)|$ and $\phi(f)$ are the magnitude and the phase of the theoretical transfer function [Eq. (30)] and $\langle\ \rangle$ denotes the average value.

A normalized propagation distance is used here. It is the ratio between the distance Δz and the base distance $\Delta z_0 = c\Delta t = c/f_e$. Both errors in the magnitude and the phase of the transfer function are computed when $\Delta z/\Delta z_0 = \Delta z/c\Delta t$ varies.

1. Influence of the propagation distance on the frequency behavior of the impulse response

The results highlighting the variations in the magnitude and the phase errors as functions of $\Delta z/\Delta z_0$ are given in Fig. 4 for two different transition frequencies: $f_r = 600$ Hz and $f_r = 1000$ Hz. For each normalized distance $\Delta z/\Delta z_0$, the errors are computed according to a transfer function obtained by the discrete Fourier transform of the analytical impulse response sampled at $f_e = 16\,000$ Hz. In this case, $\Delta z_0 = c\Delta t = 0.0215$ m. In Fig. 4(a), for both values of f_r, the error in the magnitude of the transfer function linearly increases as a function of $\Delta z/\Delta z_0$. In Fig. 4(b), for $f_r = 600$ Hz and $f_r = 1000$ Hz, the error in the phase of the transfer function is increasing toward a stable value for $\Delta z/\Delta z_0$ (toward $\Delta z/\Delta z_0 \approx 5$ for $f_r = 1000$ Hz and toward $\Delta z/\Delta z_0 \approx 9$ for $f_r = 600$ Hz) and then remains largely unchanged. The errors in the magnitude and the phase of the transfer function related to the forward radiating problem both increase when the reconstruction plane moves away from the acoustic sources.

It is also noticeable in Fig. 4 that for a given value of $\Delta z/\Delta z_0$, the errors in the magnitude and the phase are higher for $f_r = 1000$ Hz than for $f_r = 600$ Hz. In the next part, the influence of the transition frequency f_r on the frequency behavior of the impulse response is investigated.

2. Influence of the transition frequency on the frequency behavior of the impulse response

The influence of the variations of the transition frequency on the spectral behavior of the impulse response $h(\Omega_r, \tau, t)$ is highlighted in Fig. 5 where the errors in the magnitude and in the phase are represented as functions of f_r for two sampling frequencies applied to $h(\Omega_r, \tau, t)$: $f_e = 8000$ Hz and $f_e = 16\,000$ Hz. Here $\Delta z = 0.043$ m; hence for $f_e = 8000$ Hz, $\Delta z/\Delta z_0 = \Delta z/c\Delta t = 1$ and for $f_e = 16\,000$ Hz, $\Delta z/c\Delta t = 2$. Figure 5 clearly shows that the errors computed in the magnitude and phase both increase as a function of the transition frequency. For $f_e = 16\,000$ Hz, the error in the magnitude in Fig. 5(a) increases almost linearly but remains relatively weak. For $f_e = 8000$ Hz, the error is weak below $f_r = 2400$ Hz but increases very strongly for transition frequencies above. For lower sampling frequencies, the error in the magnitude of the transfer function is more sensitive to the increase in the transition frequency. There is also the same tendency for the error in the phase [see Fig. 5(b)]. The variations in the curves are similar for both sampling frequencies; however, for a given transition frequency, the error is greater for $f_e = 8000$ Hz than for $f_e = 16\,000$ Hz.

For high transition frequencies, the spectral behavior of the impulse response diverges from the theoretical model. Thus, when computing the radiated pressure field, some distorsions may appear due to the convolution between the in-

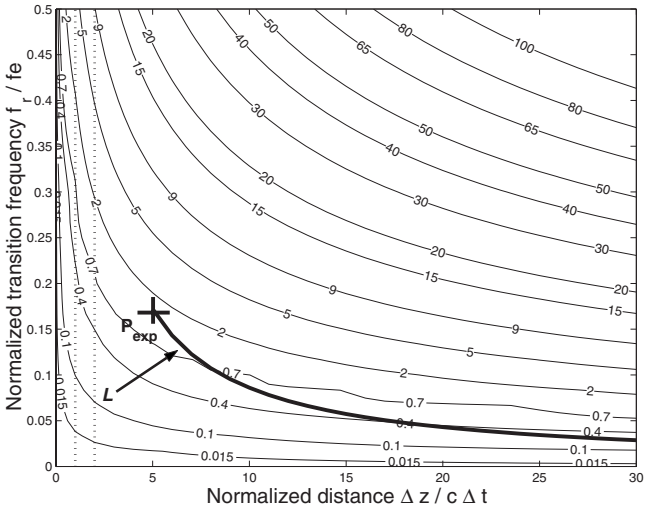


FIG. 6. Error $E_{|H_F|}$ in the magnitude of the transfer function $H_F(f)$ in the plane $(\Delta z/c\Delta t, f_r/f_e)$ and curve (L) giving $f_{r_{\max}}/f_e$ as a function of $\Delta z/c\Delta t$. The configuration parameters are $\Delta z=0.1075$ m, $\Delta L=0.0625$ m, and $f_{r_{\max}}=c/2\Delta L=2752$ Hz. The vertical lines passing by the points $\Delta z/c\Delta t=2$ and $\Delta z/c\Delta t=1$ are shown as dotted lines.

stantaneous wavenumbers resulting from the measurements and the sampled impulse response $h(\Omega_r,\tau,t)$ particularly for large wavenumbers.

3. Influence of the sampling frequency on the frequency behavior of the impulse response

The errors in the magnitude and the phase of the transfer function are now given in a plane $(\Delta z/c\Delta t, f_r/f_e)$ for different sampling frequencies in order to explain the curves in Figs. 4 and 5. Whatever the sampling frequency is chosen, typical maps highlighting isovalue lines are obtained for the error in the magnitude (see Fig. 6) and for the error in the phase (see Fig. 7). It is noticeable that the choice of two normalized values f_r/f_e and $\Delta z/\Delta z_0=\Delta z/c\Delta t$ leads to the same errors in both magnitude and phase of the transfer function computed by the Fourier transform of the sampled impulse response whatever the sampling rate f_e used. 256 samples are considered for the impulse response.

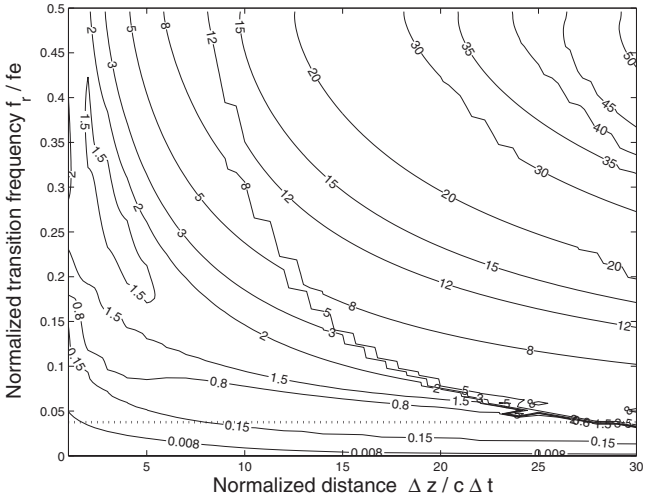


FIG. 7. Error E_{ϕ_F} (rad) in the phase of the transfer function $H_F(f)$ in the plane $(\Delta z/c\Delta t, f_r/f_e)$. The horizontal line passing by the point $f_r/f_e=0.0375$ is shown as a dotted line.

The tendency of the phase error curve in Fig. 4(b) obtained for $f_r=600$ Hz can be explained by drawing in Fig. 7 a virtual horizontal straight line passing by the point $f_r/f_e=600/16\,000=0.0375$. When $\Delta z/\Delta z_0$ varies from 0 to 15, the line first crosses several isovalues of the error, which justifies the increasing part of the curve in Fig. 4(b), and then the line becomes tangent to an isovalue, which explains the constant behavior from $\Delta z/\Delta z_0\approx 9$ in Fig. 4(b). The plot of the error in the magnitude in Fig. 6 helps to explain the variations in Fig. 5(a). Indeed by drawing a virtual vertical straight line passing by the point $\Delta z/c\Delta t=2$ in Fig. 6, one can read the values taken by the curve in Fig. 5 for $f_e=16\,000$ Hz. These values are provided by the intersection between the vertical straight line and the isovalues of the error from $f_r/f_e=0$ to $f_r/f_e=4000/16\,000=0.25$. The curve in Fig. 5 for $f_e=8000$ Hz is given by the intersection in Fig. 6 between the vertical straight line passing through point $\Delta z/c\Delta t=1$ (drawn from $f_r/f_e=0$ to $f_r/f_e=4000/8000=0.5$) and the isovalue errors.

The aim of the study is now to investigate the influence of an increase in the sampling frequency for a given numerical setup. For the acquisition stage, the step size in both x and y directions is $\Delta L=0.0625$ m. The sampling frequency is $f_e=16\,000$ Hz and the propagation distance is set to $\Delta z=0.1075$ m. The maximum transition frequency allowed which fulfills the Shannon condition of space sampling is

$$f_{r_{\max}}=\frac{ck_{\max}}{2\pi}, \tag{33}$$

with $k_{\max}=\pi/\Delta L$, the wavenumber limit.

Finally $f_{r_{\max}}$ can be written as

$$f_{r_{\max}}=\frac{c}{2\Delta L}. \tag{34}$$

It is also true that the maximum transition frequency must fulfill the Shannon condition in the time domain. In fact, the smallest value of the couple of values $(f_e/2, c/2\Delta L)$ must be chosen for $f_{r_{\max}}$. For this numerical setup, $f_{r_{\max}}=\min(8000\text{ Hz}, 2752\text{ Hz})=2752$ Hz.

In this configuration, the transition frequency varies from 0 to $f_{r_{\max}}=c/2\Delta L=2752$ Hz. Thus, if f_e is not lower, the error values in the area $f_r/f_e>f_{r_{\max}}/f_e$ in Figs. 6 and 7 will never be reached. It is also interesting to observe the values of the error when the sampling frequency increases.

$f_{r_{\max}}/f_e$ can be written as

$$\frac{f_{r_{\max}}}{f_e}=\frac{C_{\exp}}{\Delta z/c\Delta t}, \tag{35}$$

where $C_{\exp}=f_{r_{\max}}\Delta z/c$ is a constant term depending on the numerical or experimental setup. Here $C_{\exp}=\Delta z/2\Delta L=0.86$. Thus, the values taken by $f_{r_{\max}}/f_e$ are given by $0.86/(\Delta z/c\Delta t)$. The numerical setup ($f_e=1/\Delta t=16\,000$ Hz, $\Delta z=0.1075$ m, $\Delta L=0.0625$ m) provides the coordinates of $P_{\exp}$ (see Fig. 6) the first point of the curve L defined by $f_{r_{\max}}/f_e=0.86/(\Delta z/c\Delta t)$ in the plane $(\Delta z/c\Delta t, f_r/f_e)$: $P_{\exp}(\Delta z/c\Delta t, f_{r_{\max}}/f_e)=P_{\exp}(5, 0.172)$. When f_e is increased

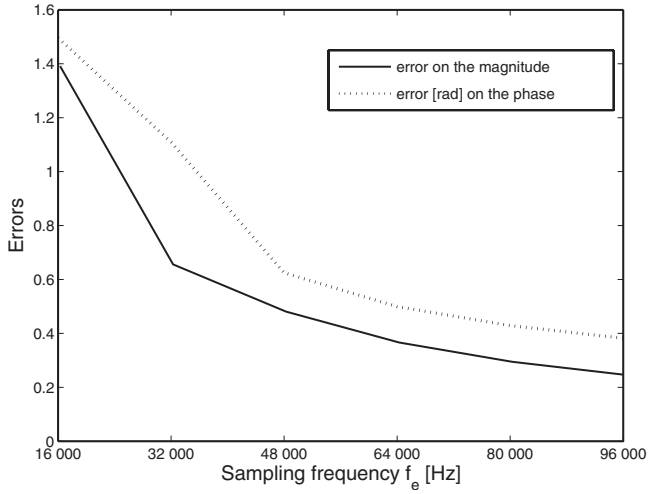


FIG. 8. Errors in the magnitude and in the phase (rad) of the transfer function $H_F(f)$ as functions of the sampling frequency f_e for the transition frequency f_r=2752 Hz and the propagation distance Δz=0.1075 m.

from 16 000 to 96 000 Hz, the curve L decreases and crosses the isovalue lines toward the decreasing error values. It can be concluded that an increase in the sampling frequency causes the errors to decrease for each impulse response $h(\Omega_r,\tau,t)$ of the filters used in the time-wavenumber domain to solve the radiation problem. The sampling frequency can be increased at the acquisition stage or *a posteriori* with a Shannon interpolation of the sampled signal.

The reduction in the errors when the sampling frequency increases is shown in Fig. 8, where the errors in both magnitude and phase are obtained from the previous numerical setup: Δz=0.1075 m and f_r=2752 Hz. These values are collected from the isovalue lines crossed by the curve L in Fig. 6.

The fact that an increase in the sampling frequency leads to a reduction in the errors in both magnitude and phase of the transfer function has been highlighted. However, even if the use of a high sampling frequency seems necessary, it is not sufficient, which is illustrated in Fig. 9. In this figure, the

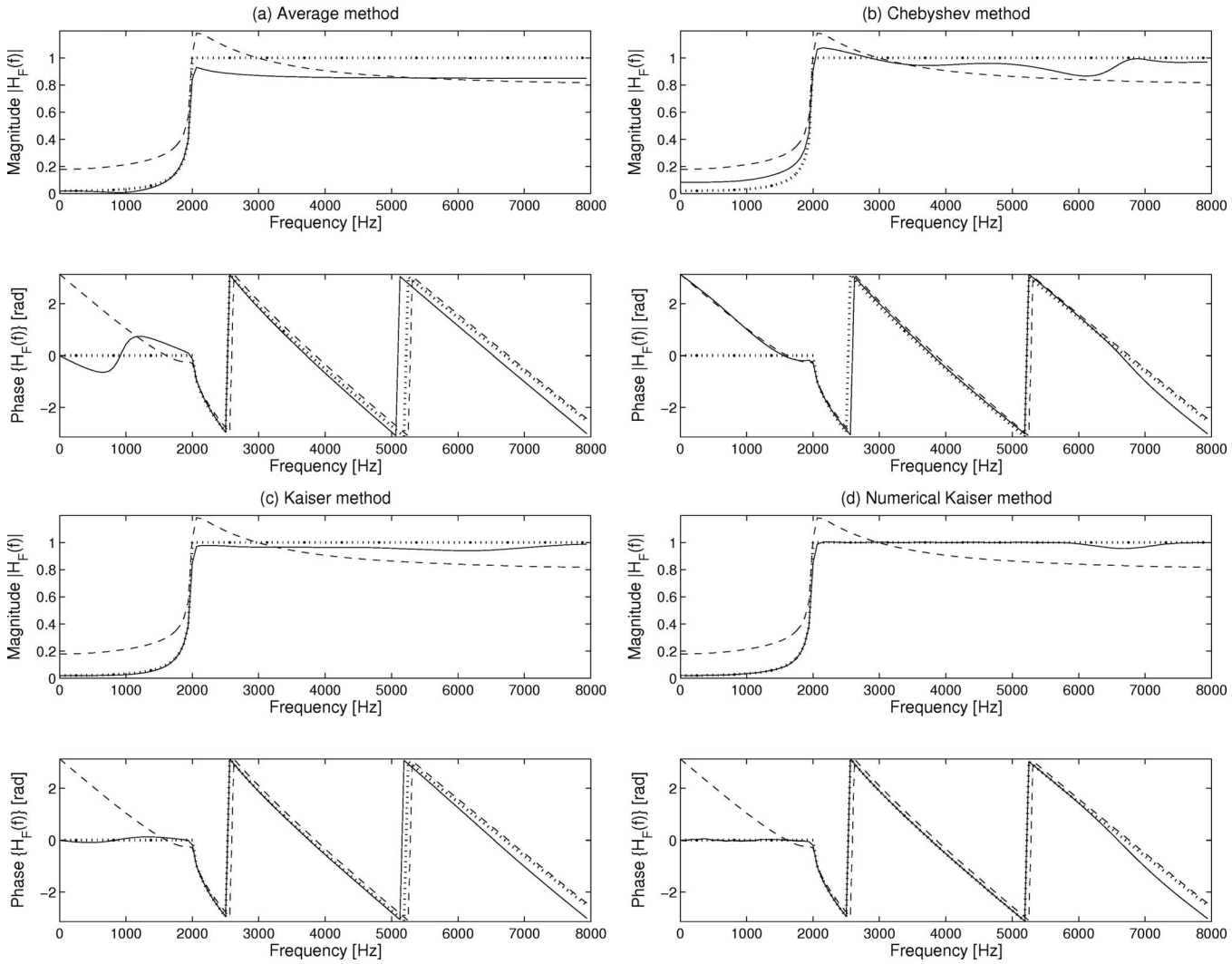


FIG. 9. Transfer functions $H_F(\Omega_r,\tau,f)$ shown as solid lines (magnitude and phase in rad) after processing $h(\Omega_r,\tau,t)$. N=256 points, f_e=16 000 Hz, Δz =0.1075 m ($\Delta z/c\Delta t$=5), and f_r=2000 Hz (f_r/f_e=0.125). The theoretical transfer function $H(\Omega_r,\tau,f)$ is indicated as a dotted line. The Fourier transform of $h(\Omega_r,\tau,t)$ sampled at f_e=64 000 Hz is shown as a dashed line. (a) Sampling h using the average method, (b) Chebyshev method (cutoff frequency f_c =6400 Hz and upsampling factor of D=8), (c) low-pass filtering using a Kaiser–Bessel filter (f_c=6640 Hz and D=2), and (d) numerical Kaiser method (f_c=6780 Hz).

transfer function (magnitude and phase) $H_F(\Omega_r,\tau,f)$ computed by Fourier transform of $h(\Omega_r,\tau,t)$ is drawn as a dashed line for $f_e=64\,000$ Hz, $\Delta z=0.1075$ m [i.e., $(\Delta z/c\Delta t)=20$] and $f_r=2000$ Hz [i.e., $f_r/f_e=0.031\,25$]. For these two values $(\Delta z/c\Delta t, f_r/f_e)$, the error in the magnitude is $E_{|H_F|}=0.2$ and the error in the phase is $E_{\phi_F}=0.45$ (the values can also be extracted from Figs. 6 and 7). The use of a high sampling frequency applied to the impulse response is advantageous because $E_{|H_F|}=0.6$ and $E_{\phi_F}=1.3$ for $f_e=16\,000$ Hz. However, the result is not satisfactory, as shown in Fig. 9, where the transfer function, shown as a dashed line, is far from the theoretical model, shown as a dotted line. This result leads us to consider additional processing of the impulse response in order to provide the pressure field radiated by the acoustic sources on a forward plane. These processing techniques are detailed in Sec. IV B.

B. Processing for providing an operational impulse response

Two approaches are considered here. The first is based on the analytical formulation of the impulse response $h(\Omega_r,\tau,t)$ given in Eq. (26) in the time-wavenumber domain. The second starts from the theoretical frequency response $H(\Omega_r,\tau,\omega)$ in Eq. (30) and by using an inverse Fourier transform yields the impulse response.

1. Processing from the analytical impulse response

For this case, consider the following equation derived from Eq. (26) giving the impulse response:

$$h(\Omega_r,\tau,t)=\delta(t-\tau)-g(\Omega_r,\tau,t), \tag{36}$$

where

$$g(\Omega_r,\tau,t)=\pi\Omega_r^2\frac{J_1(\Omega_r\sqrt{t^2-\tau^2})}{\Omega_r\sqrt{t^2-\tau^2}}\Gamma(t-\tau). \tag{37}$$

As sampling the impulse response even with a relatively high rate may lead to distorsions in the transfer function, direct sampling is replaced by average sampling. Instead of considering $g[n]$, the sampling value of $g(t)$ at the time $t=n\Delta t$, the average value $\bar{g}[n]$ is computed into an interval Δt centered at $t=n\Delta t$:

$$\bar{g}[n]=\frac{1}{\Delta t}\int_{n\Delta t-(\Delta t/2)}^{n\Delta t+(\Delta t/2)}g(t)dt. \tag{38}$$

The integral in Eq. (38) is approximated by the trapezoidal formula.

Another modification is used to overcome the problem of the impulse response whose transfer function is not band limited. The process consists first of increasing the sampling rate of the impulse response by a factor D so that the new sampling frequency is $f_e'=Df_e$. The response contains DN samples. Then the upsampled response is filtered using a low-pass filter. Finally, the filtered impulse response is downsampling by the factor $1/D$ to ensure that the final sampling frequency f_e'/D matches f_e, that of the acoustic signals acquired. The number of samples of the resulting impulse response is N.

The use of two low-pass filters have been investigated, one with an infinite impulse response, a Chebyshev filter, and the other one with a finite impulse response (FIR) given by associating a cardinal sine with a Kaiser–Bessel window.[16] The choice of a Chebyshev filter which exhibits equiripple behavior in the passband and a monotonic characteristic in the stopband facilitates the implementation. The advantage of the FIR filter is that it has a linear-phase characteristic within the passband. Then it is easy to postprocess the filtered impulse response to recover the phase. In addition, the Kaiser–Bessel window which decays toward zero gradually permits to alleviate the presence of large oscillations in both the passband and the stopband of the frequency response.

The impulse response of the FIR filter is

$$w(t)=\frac{I_0(\beta\sqrt{1-(2t/T)^2})}{I_0(\beta)}\frac{\sin(\alpha\pi tf_e)}{\pi t}. \tag{39}$$

I_0 is the modified Bessel function of the first kind and order 0. T is the duration of the Kaiser–Bessel window. α is linked to the cutoff frequency f_c of the low-pass filter and β is a shape parameter of the Kaiser–Bessel window:

$$\alpha=\frac{2f_c}{f_e}, \tag{40}$$

$$\beta\approx 0.1102(A-8.7), \tag{41}$$

where A is the sidelobe attenuation in decibels.

Two ways of implementing the convolution between the impulse response and the low-pass filter can be considered. First, the filtered response $g_f(t)$ can be provided using a discrete sum as

$$g_f[n]=\sum_m w[m]g[n-m]. \tag{42}$$

But it can be also computed using a numerical approximation of the following integral given by the trapezoidal method:

$$g_f(t)=\int_{t-T/2}^{t+T/2}g(\theta)w(t-\theta)d\theta. \tag{43}$$

2. Processing from the theoretical frequency response

The Fourier transform of Eq. (36) yields

$$H(\Omega_r,\tau,\omega)=e^{-j\omega\tau}-G(\Omega_r,\tau,\omega). \tag{44}$$

Since $H(\Omega_r,\tau,\omega)$ is analytically defined in Eq. (30), so it is for the transfer function $G(\Omega_r,\tau,\omega)$ whose both theoretical magnitude and phase are easily deduced.

It follows

$$G(\Omega_r,\tau,\omega)=e^{-j\omega\tau}-H(\Omega_r,\tau,\omega). \tag{45}$$

By applying an inverse Fourier transform either to $H(\Omega_r,\tau,\omega)$ [Eq. (30)] or to $G(\Omega_r,\tau,\omega)$ [Eq. (45)], the im-

pulse response $h(\Omega_r,\tau,t)$ or $g(\Omega_r,\tau,t)$ is obtained. From Eq. (36), both approaches provide finally the same impulse response $h(\Omega_r,\tau,t)$.

C. Comparisons between transfer functions resulting from processing

The aim of this part is to compare several transfer functions $G(\Omega_r,\tau,\omega)$ resulting from different processing techniques in order to evaluate their effectiveness in resolving the source radiating problem. Four treatments are applied to the theoretical function $g(\Omega_r,\tau,t)$ in Eq. (37). They are summarized as follows.

- *Average method.* $g(\Omega_r,\tau,t)$ is average sampled according to Eq. (38).
- *Chebyshev method.* $g(\Omega_r,\tau,t)$ is low-pass filtered using a Chebyshev filter with a cutoff frequency $f_c=6400$ Hz. It is achieved by upsampling $g(\Omega_r,\tau,t)$ by the factor $D=8$ using the low-pass filter and then downsampling the resulting response by the factor $1/D$.
- *Kaiser method.* $g(\Omega_r,\tau,t)$ is average sampled and low-pass filtered using a Kaiser–Bessel filter with a cutoff frequency $f_c=6640$ Hz. An upsampling factor of $D=2$ is used.
- *Numerical Kaiser method.* The same Kaiser–Bessel filter is applied but the integral in Eq. (43) is numerically computed using the trapezoidal method.

For all cases, $g(\Omega_r,\tau,t)$ is initially sampled with the sampling frequency $f_e=16\,000$ Hz giving 256 samples. The propagation distance and the transition frequency are set to $\Delta z=0.1075$ m and $f_r=2000$ Hz. Figure 9 highlights the transfer functions $H(\Omega_r,\tau,f)$ (magnitude and phase) for the four different processing techniques. The frequency responses are obtained by applying a Fourier transform to the sampled response $g(\Omega_r,\tau,t)$ before using Eq. (44).

By comparing the four transfer functions to the dashed line (see Fig. 9), it seems evident that the transfer functions provided by processing $g(\Omega_r,\tau,t)$ are more relevant than the one obtained by operating a Fourier transform directly on the sampled response even though the sampling frequency is higher (64 000 Hz instead of 16 000 Hz). In addition, filtering $g(\Omega_r,\tau,t)$ in order to limit its frequency band is advantageous. The use of average sampling with no filter is less effective than filtering, in particular, in the frequency area of propagating components. The use of a FIR filter with a Kaiser–Bessel window seems more accurate than the use of a Chebyshev filter especially for the phase. The most accurate transfer function in Fig. 9 is obtained by numerically computing the integral of convolution involving the Kaiser–Bessel filter. One can note that this comparison must be done for each transition frequency and then for each point of the wavenumber domain.

These remarks can also be verified by considering the variations in the magnitude and phase errors of Eqs. (31) and (32) when the transfer functions $H_F(f)$ are computed from the Chebyshev, the average, the Kaiser, and the numerical Kaiser methods. For this purpose, Figs. 10 and 11 are to be compared with Figs. 4 and 5. It is clear that processing the impulse response leads to a reduction in the errors. The lowest errors are achieved by using the numerical Kaiser method while the Chebyshev method seems to generate phase errors (see Fig. 10) when the propagation distance increases. Increasing the sampling rate of the impulse response by a factor higher than 8 could enhance the performance of the Chebyshev method.

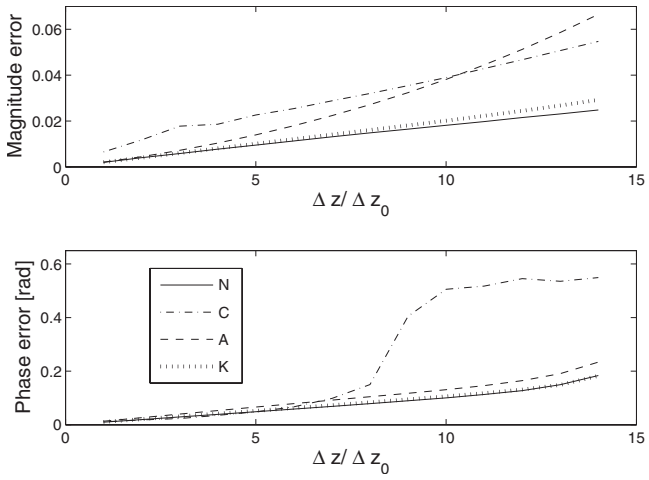


FIG. 10. Errors in the magnitude $E_{|H_F|}$ and in the phase E_{ϕ_F} (rad) of the transfer function $H_F(f)$ as functions of the normalized propagation distance $\Delta z/\Delta z_0$ with $\Delta z_0=c\Delta t=0.0215$ m. The normalized transition frequency is $f_r=1000$ Hz. The transfer functions are obtained by applying the Fourier transform to the impulse responses computed by Chebyshev (C), average (A), Kaiser (K) and numerical Kaiser (N) methods for $f_e=16\,000$ Hz.

It is of course true that the best matching between the theoretical and the computed transfer functions occurs for the method based on the inverse Fourier transform (see Sec. IV B 2), which is evident as the starting point of the approach, called here Fourier method, is precisely the theoretical frequency response.

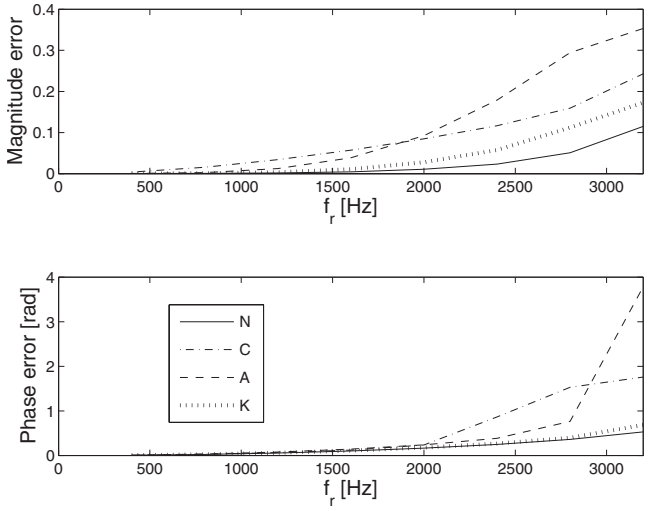


FIG. 11. Errors in the magnitude $E_{|H_F|}$ and in the phase E_{ϕ_F} (rad) of the transfer functions $H_F(f)$ as functions of the transition frequency ($\Delta z=0.043$ m) for $f_e=8000$ Hz ($\Delta z/c\Delta t=1$). The transfer functions are obtained by applying the Fourier transform to the impulse responses computed by Chebyshev (C), average (A), Kaiser (K), and numerical Kaiser (N) methods.

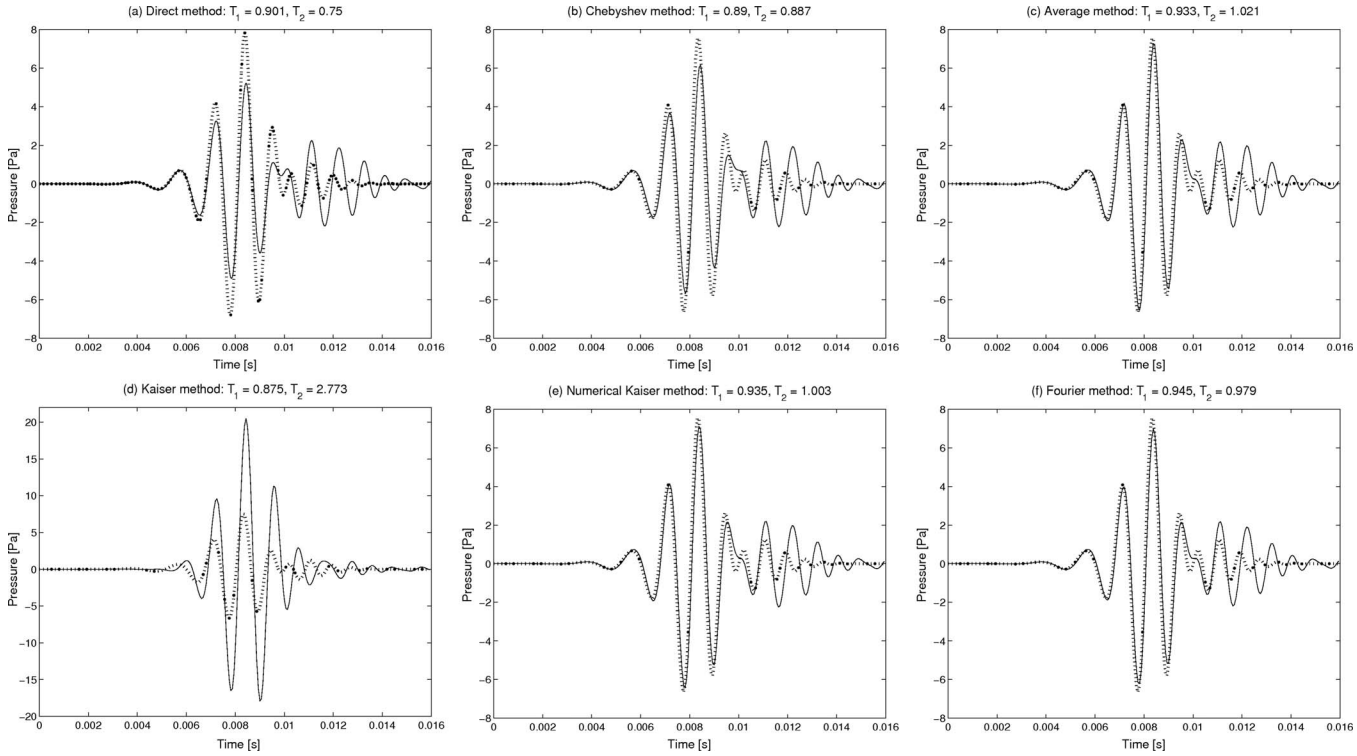


FIG. 12. Reconstructed temporal signals (solid line) versus reference signals (dotted line) in the time-space domain in location P_3 (see Fig. 1) for different impulse responses computed by six methods [direct (a), Chebyshev (b), average (c), Kaiser (d), numerical Kaiser (e), and Fourier (f)]. The indicators T_1 [Eq. (47)] and T_2 [Eq. (48)] are given on top of each graph.

V. NUMERICAL RESULTS

A. Setup

The source plane is composed of three monopoles at the positions M_1 (0.3125 m, 0.375 m, 0 m), M_2 (0.75 m, 0.75 m, 0 m), and M_3 (0.25 m, 0.75 m, 0 m). Both monopoles M_1 and M_2 generate a signal with a linear frequency modulation in the band [200 Hz, 1800 Hz] and a Gaussian amplitude modulation while monopole M_3 radiates a Morlet wavelet whose expression is

$$s(t) = \cos(2\pi f_0 t)e^{-t^2/2}, \tag{46}$$

with $f_0=800$ Hz. Thus the sources are nonstationary. The simulation of the acquisition is done using a 17×17 microphone array located in the measurement plane $z=z_A$ with $z_A=0.05$ m. The step size in both x and y directions is $\Delta L=0.0625$ m, providing an overall aperture size of 1.0×1.0 m^2. The propagation distance is $\Delta z=0.1075$ m. Thus, the forward plane is located at $z_F=0.1575$ m, as shown in Fig. 1. The emitted signals are sampled at a frequency of $f_e=16\,000$ Hz providing 256 samples.

The aim of this study is to reconstruct the time evolving pressure field at each point of the forward plane in front of the square grid of 17×17 virtual microphones using Eq. (25). Five different impulse responses $h(\Omega_r,\tau,t)$ in the time-wavenumber domain are investigated. They are computed from the Chebyshev method, the Kaiser method, the numerical Kaiser method, the Fourier method, and the direct method for which the impulse response $h(\Omega_r,\tau,t)$ is provided by directly sampling $g(\Omega_r,\tau,t)$ in Eq. (37) at $f_e=64\,000$ Hz.

B. Indicators

In order to comment the results obtained objectively, two temporal indicators T_1 and T_2 are proposed [see Eqs. (47) and (48)]. They are based on the reconstructed signals $p(x,y,z_F,t)$ but also on simulated signals $p_r(x,y,z_F,t)$ directly propagated on the forward plane $z=z_F$, which are considered as reference signals:

$$T_1 = \frac{\langle p_r(x,y,z_F,t)p(x,y,z_F,t)\rangle}{\sqrt{\langle p_r^2(x,y,z_F,t)\rangle\langle p^2(x,y,z_F,t)\rangle}}, \tag{47}$$

$$T_2 = \sqrt{\frac{\langle p^2(x,y,z_F,t)\rangle}{\langle p_r^2(x,y,z_F,t)\rangle}}. \tag{48}$$

$\langle\ \rangle$ is the average value.

T_1 is a correlation coefficient which is sensitive to the similarity between the shapes of the signals and thus between their phase difference. T_2 is the ratio between two root mean square values for characterizing the similarity of the amplitudes of both signals.

C. Results in the time-space domain

Figure 12 highlights the temporal pressure signals $p(0.25\text{ m},0.75\text{ m},0.1575\text{ m},t)$ radiated in P_3 whose location is indicated in Fig. 1. P_3 is in front of the monopoles M_3. The pressure signals are provided by the method proposed using the five different impulse responses in the time-wavenumber domain mentioned above plus the impulse response obtained by the inverse Fourier transform. They are compared to the reference pressure signals directly propagated to the forward

plane $z=z_F$ by simulation. The indicator values T_1 and T_2 are reported in Table I. The examination of both Fig. 12 and Table I leads to the ranking of the methods in order of increasing relevance: Kaiser method (K), direct method (D), average method (A), Chebyshev method (C), numerical Kaiser method (N), and Fourier method (F). The Kaiser method seems to suffer from a two weak oversampling factor of 2. Therefore, the use of a low-pass Kaiser–Bessel filter with a numerical implementation or the inverse Fourier transform of the analytical transfer function leads to the most operational impulse response in the time-wavenumber domain. In addition, the Fourier method is more efficient than the numerical Kaiser method.

Results are enhanced when the impulse response $h(\Omega_r,\tau,t)$ is low-pass filtered. It confirms the fact that $g(\Omega_r,\tau,t)$ in Eq. (37) should not only be sampled. With filtering or using the Fourier transform, results are improved in the whole space, which can be verified by comparing the

TABLE I. Indicators T_1 [Eq. (47)] and T_2 [Eq. (48)] computed from reference signals and pressure signals computed in $z=z_F$ from impulse responses obtained by the direct method (D), the Kaiser method (K), the average method (A), the Chebyshev method (C), the numerical Kaiser method (N), and the Fourier method (F).

				P_2		
	D	K	A	C	N	F
T_1	0.929	0.944	0.958	0.920	0.961	0.963
T_2	0.812	2.397	1.054	0.951	1.009	0.971
				P_3		
	D	K	A	C	N	F
T_1	0.901	0.875	0.933	0.890	0.935	0.945
T_2	0.750	2.773	1.021	0.887	1.003	0.979
				P_4		
	D	K	A	C	N	F
T_1	0.991	0.689	0.989	0.990	0.989	0.991
T_2	1.022	1.025	1.056	1.043	1.063	1.063

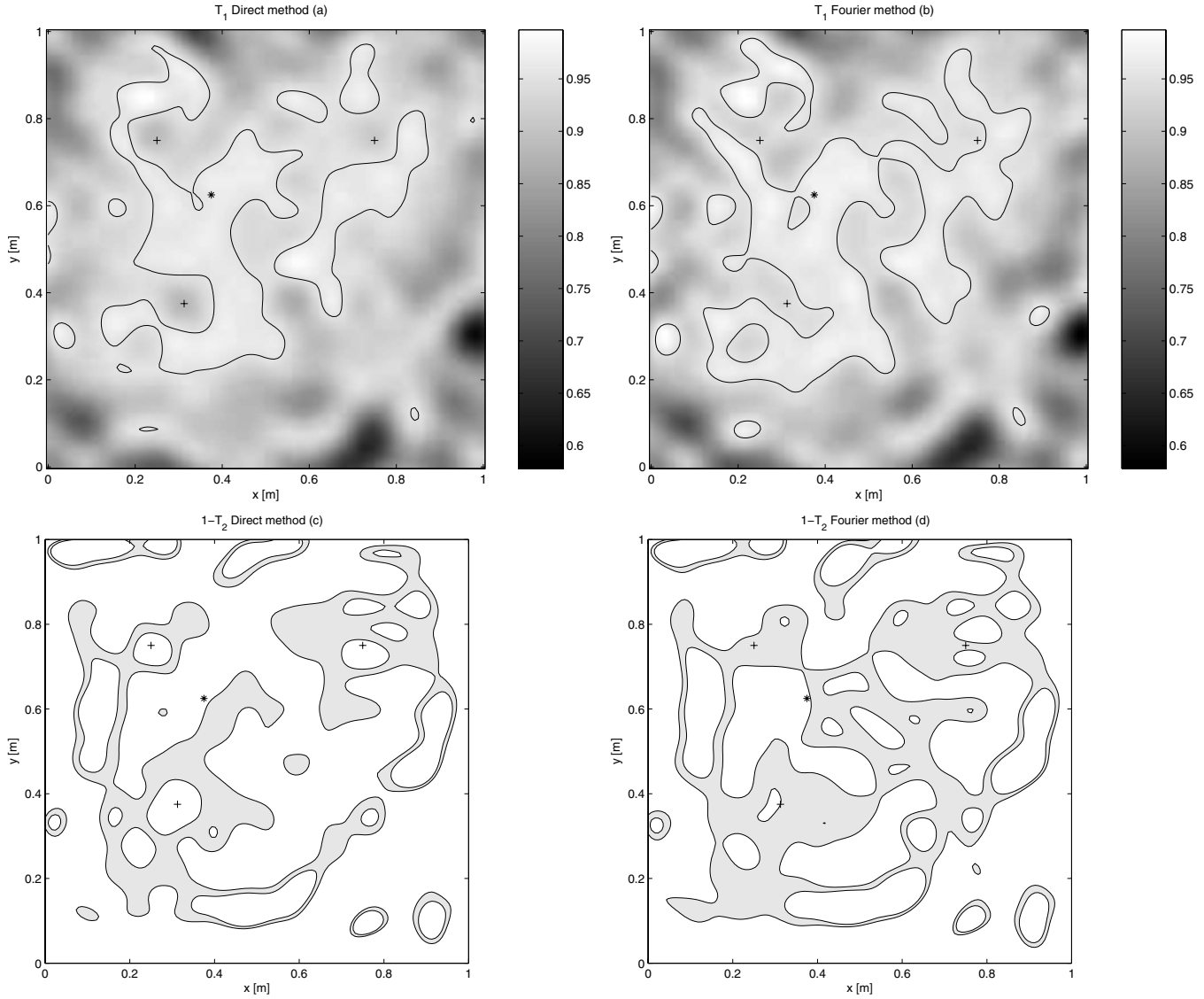


FIG. 13. Spatial maps of indicator T_1 and $1-T_2$ to assess the phase and the amplitude similarities between the reference signals and the projected signals using the direct method [(a) and (c)] with the sampling frequency $f_e=64\,000$ Hz and the Fourier method [(b) and (d)] with $f_e=16\,000$ Hz. The locations of $P_1(+)$, $P_2(+)$, $P_3(+)$, and $P_4(*)$ are marked. [(a) and (c)] The contour line is at the value 0.95. [(c) and (d)] The areas in gray correspond to values of $1-T_2$ within the interval $[-0.05,0.05]$.

reconstructed temporal signals from other locations of the space with the reference signals as in Fig. 12. For this purpose, spatial cartographies are reported for the direct method and the Fourier method in Fig. 13. The 0.95 contour line is displayed for indicator T_1 in Figs. 13(a) and 13(b). One can observe that the locations facing the monopoles do not give the best results in terms of phase difference. However, the error obtained by directly sampling the impulse response with $f_e=64\,000$ Hz is decreased when the Fourier method is applied even with a lower sampling frequency ($f_e=16\,000$ Hz). The error is also higher near the edges of the scanned area. In Figs. 13(c) and 13(d), the spatial map is given for indicator $1-T_2$. Indeed the similarity between the amplitudes of the reference signals and the projected signals is evident when $1-T_2$ is close to 0. In fact, the lowest error is reached for spatial locations within the area in gray for $1-T_2$ values in the interval $[-0.05,0.05]$. It is noticeable that this area is larger for the Fourier method than from the direct method. In most locations, the amplitudes of the projected signals are overestimated. However, the amplitudes of the projected signals are inclined to be underestimated by the direct method for locations around the source monopoles. The Fourier method gives better results near the monopoles. In the same way as the spatial map for indicator T_1, the error for $1-T_2$ is higher near the border of the antenna.

VI. CONCLUSION

The method proposed to forward propagate time-evolving acoustic fields has the singularity to be based on a convolution product in the time-wavenumber domain involving an analytical impulse response. However, this impulse response needs to be carefully implemented. Indeed, it is shown here that just sampling the response is not sufficient to deduce the time-dependent pressure field radiated by the sources. Two processing methods give accurate results. The first method applies a Kaiser–Bessel low-pass filter to the impulse response which is average sampled and the convolution integral is numerically computed using the trapezoidal technique. The second method provides the impulse response by computing an inverse Fourier transform of the analytical transfer function.

The errors made on the propagated acoustic fields increase with the distance separating the sources from the reconstructed plane and also with the increase in a specific parameter, the transition frequency which depends on each wavenumber and separates two tendencies of the travelling waves: propagating waves and evanescent waves. It is clear that upsampling the impulse response reduces the errors. It is remarkable that whatever the sampling frequency, the propagation distance, or the transition frequency chosen, the errors are the same as long as the normalized frequency (of the transition frequency by the sampling frequency) and the normalized distance (of the propagation distance by the distance covered during a sampling step) remain constant.

The main interest of the method is that it provides the instantaneous acoustic field radiated from the sources. The approach could be applied to active control and to diagnose and monitor various acoustic systems and to calculate the radiation from structures under unsteady excitations such as those produced by impacts or turbulent flows.

[1] S. G. Hill, S. D. Snyder, and B. S. Cazzolato, "Deriving time-domain models of structural-acoustic radiation into free space," Mech. Syst. Signal Process. **19**, 1015–1033 (2005).

[2] S. J. Elliott and M. E. Johnson, "Radiation modes and the active control of sound power," J. Acoust. Soc. Am. **94**, 2194–2204 (1993).

[3] A. Chaigne and C. Lambourg, "Time-domain simulation of damped impacted plates. i. theory and experiments," J. Acoust. Soc. Am. **109**, 1422–1432 (2001).

[4] G. Derveaux, A. Chaigne, P. Joly, and E. Bécache, "Time-domain simulation of a guitar: model and method," J. Acoust. Soc. Am. **114**, 3368–3383 (2003).

[5] A. Akay and M. Latcha, "Sound radiation from an impact-excited clamped circular plate in an infinite baffle," J. Acoust. Soc. Am. **74**, 640–648 (1983).

[6] P. Troccaz, R. Woodcock, and F. Laville, "Acoustic radiation due to the inelastic impact of a sphere on a rectangular plate," J. Acoust. Soc. Am. **108**, 2197–2202 (2000).

[7] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, New York, 1999).

[8] J.-H. Thomas, V. Grulier, S. Paillasseur, J.-C. Pascal, and J.-C. Le Roux, "Real-time nearfield acoustic holography (RT-NAH): A technique for time-continuous reconstruction of a source signal," in Proceedings of Novem 2005, Saint-Raphaël, France (2005), Paper No. 139.

[9] M. C. Junger and D. Feit, *Sound, Structures, and Their Interaction* (MIT, Cambridge, MA, 1986) (republication by the Acoustical Society of America, 1993).

[10] E. G. Williams and J. D. Maynard, "Numerical evaluation of the rayleigh integral for planar radiators using the FFT," J. Acoust. Soc. Am. **72**, 2020–2030 (1982).

[11] O. de la Rochefoucauld, M. Melon, and A. Garcia, "Time domain holography: Forward projection of simulated and measured sound pressure fields," J. Acoust. Soc. Am. **116**, 142–153 (2004).

[12] M. Forbes, S. Letcher, and P. Stepanishen, "A wave vector, time-domain method of forward projecting time-dependent pressure fields," J. Acoust. Soc. Am. **90**, 2782–2792 (1991).

[13] V. Grulier, "Propagation directe et inverse dans l'espace temps-nombre d'onde: Application à une méthode d'holographie acoustique de champ proche pour les sources non stationnaires (Direct and inverse propagation in the time-wavenumber domain: Application to a nearfield acoustic holography method for non stationary sources)," Thesis of the Université du Maine, Le Mans, France (2005).

[14] V. Grulier, J.-H. Thomas, J.-C. Pascal, and J.-C. Le Roux, "Time varying forward projection using wavenumber formulation," Proceedings of Internoise2004, Prague, Czech Republic (2004), Paper No. 406.

[15] J. Hladik, *La transformation de Laplace à plusieurs variables (The Multi-Variable Laplace Transform)* (Masson, Paris, 1969).

[16] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed. (Prentice-Hall, Upper Saddle River, NJ, 1999).

Analysis and design of gammatone signal models

Stefan Strahl[a)]
International Graduate School for Neurosensory Science and Systems, Carl von Ossietzky University, D-26111 Oldenburg, Germany

Alfred Mertins
Institute for Signal Processing, University of Lübeck, Ratzeburger Allee 160, D-23538 Lübeck, Germany

(Received 14 January 2009; revised 22 June 2009; accepted 29 July 2009)

An established model for the signal analysis performed by the human cochlea is the overcomplete gammatone filterbank. The high correlation of this signal model with human speech and environmental sounds [E. Smith and M. Lewicki, Nature (London) **439**, 978–982 (2006)], combined with the increased time-frequency resolution of sparse overcomplete signal models, makes the overcomplete gammatone signal model favorable for signal processing applications on natural sounds. In this paper a signal-theoretic analysis of overcomplete gammatone signal models using the theory of frames and performing bifrequency analyses is given. For the number of gammatone filters $M \geq 100$ (2.4 filters per equivalent rectangular bandwidth), a near-perfect reconstruction can be achieved for the signal space of natural sounds. For signal processing applications like multi-rate coding, a signal-to-alias ratio can be used to derive decimation factors with minimal aliasing distortions.
 [DOI: 10.1121/1.3212919]

PACS number(s): 43.60.Hj [DOS] Pages: 2379–2389

I. INTRODUCTION

The earliest theoretical signal analysis model, proposed by Fourier,[1] analyzes the frequency content of a signal using the expansion of functions into a weighted sum of sinusoids. Gabor[2] extended this signal model using shifted and modulated time-frequency atoms which analyze the signal in the frequency as well as in the time dimension. With the wavelet signal model, a further improvement was presented by Morlet *et al.*[3] using time-frequency atoms that are scaled dependent on their center frequency. This yields an analysis of the time-frequency plane with a non-uniform tiling. The time-frequency atoms used in these signal models normally do not assume an underlying signal structure. As the performance of subsequent processing algorithms depends strongly on how well the fundamental features of a signal are captured, it is favorable to use time-frequency atoms that are specialized to the applied signal class. In this paper we are concerned with the signal class of natural sounds such as speech or environmental sounds, which have been found to be highly correlated with gammatone time-frequency atoms.[4,5] The signal-dependent properties of gammatone atoms are their non-uniform frequency tiling of the time-frequency plane and their asymmetric envelope.[6] A gammatone filterbank is furthermore an established model for the human auditory filters.[7–12] Several analysis-synthesis systems have been proposed using gammatone filters in the analysis and time-reversed filters in the synthesis stage,[13–16] including low-delay[17] and level-dependent asymmetric compensation[18] concepts.

Overcompleteness in signal models has advantages in signal coding applications. It enables sparse signal models like matching pursuit[19] (MP) to search for the sparsest signal representation from the resulting infinite number of possible encodings. Overcompleteness further introduces a robustness toward noise.[20,21] Generally, the choice of the number of time-frequency atoms in a signal model, hence the choice of overcompleteness, is nontrivial. In this paper we are therefore also concerned with the trade-off between the achieved performance in the subsequent processing algorithms and the introduced computational load. To derive the minimal number of time-frequency atoms needed to realize an overcomplete gammatone signal model that can adequately analyze the signal space, we use the theory of frames[22–25] which is a generalization of signal representations based on transforms and filterbanks. A second parameter that can control the overcompleteness of the gammatone signal model is the number of removed analysis filter coefficients. Such a decimation of the filter coefficients introduces aliasing distortions that should not only be kept to a minimum but should also be steered to cancel out in the synthesis stage of the filterbank. Therefore we performed a bifrequency analysis[26] in addition to a frame-theoretic analysis of overcomplete decimated gammatone signal models. We show how a signal-to-alias ratio (SAR) can be used to derive optimal sets of decimation factors with minimal aliasing distortions at a given total decimation factor.

This paper is organized as follows. In Sec. II we introduce the analyzed overcomplete gammatone signal models. In Sec. III we present a frame-theoretic analysis of a non-decimated and a decimated overcomplete gammatone signal model by performing an eigenanalysis of the frame operator.[27] We further show how these results can be used to select the optimal number of atoms for an overcomplete

[a)]Author to whom correspondence should be addressed. Electronic mail: stefan.strahl@uni-oldenburg.de

gammatone signal model. In Sec. IV we show how optimal decimation factors with minimized distortion artifacts can be derived using the bifrequency system analysis.[26] We then analyze these theoretically derived optimal parameters in Sec. V in several audio coding examples.

A. Notation

Matrices and vectors are printed in boldface. $\|\cdot\|$ denotes the Euclidean norm of a vector. $\langle\cdot,\cdot\rangle$ is the inner product of a vector space. $\mathbb{Z}$ is the set of all integers, $\mathbb{R}$ is the set of all real, and $\mathbb{C}$ is the set of all complex numbers. $[a,b] := \{x | a \leq x \leq b\}$ represents the set of all numbers between and including a and b. The superscript * denotes the complex conjugate of a complex number and the superscript ^{H} the conjugate transposition of a complex $m \times n$ matrix. The asterisk $*$ denotes convolution. The argument of the maximum of a function $f(x)$ is denoted as $\arg\max_x f(x)$.

II. OVERCOMPLETE GAMMATONE SIGNAL MODEL

A. Gammatone function

In 1960, Flanagan[28] used a gammatone function as a model of the basilar membrane displacement in the human ear. Johannesma[29] further showed in 1972 that a gammatone filter can be used to approximate responses recorded from the cochlear nucleus in the cat. In 1975, de Boer[30] used a gammatone function to model impulse responses from auditory nerve fiber recordings in the cat, which have been estimated using a linear reverse-correlation technique. The term "Gamma-tone" was introduced in 1980 by Aertsen and Johannesma.[31] Patterson *et al.*[8] stated in 1988 that the gammatone filter also delineates psychoacoustically determined auditory filters in humans. A gammatone filter is defined as

$$\gamma[n] = a n^{\nu-1} e^{-\lambda n} e^{2\pi i f_c n}, \tag{1}$$

with the amplitude a and the filter order ν. The damping factor λ is defined as $\lambda = 2\pi b \mathrm{ERB}(f_c)$ (ERB denotes equivalent rectangular bandwidth) with the center frequency f_c. The parameter b controls the bandwidth of the filter proportional to the ERB of a human auditory filter. For humans, the parameters $\nu = 4$ and $b = 1.019$ have been derived using notched-noise masking data.[32] For moderate sound pressure levels, Moore *et al.*[33] estimated the size of an ERB in the human auditory system as $\mathrm{ERB}(f_c) = 24.7 + 0.108 f_c$. The center frequencies of the gammatone filters are equally spaced on the ERB frequency scale.[34] The scale is defined as the number of ERBs below each frequency with $\mathrm{ERBS}(f_c) = 21.4 \log_{10}(0.004\,37 f_c + 1)$. This non-uniform distribution of the center frequencies (see Fig. 1) correlates with the $1/f$ distribution of frequency energy found in natural signals.[35] It is one of the signal-dependent features of a gammatone signal model. The frequency-dependent bandwidth resulting in narrower filters at low frequencies and broader filters at high frequencies is also an important feature of the gammatone time-frequency atoms. In Sec. III we will show that this enables the signal model to form a snug frame. The third signal-dependent feature of gammatone time-frequency atoms is the asymmetric envelope of the gammatone function,[6] which can also be found in natural sounds, exhibiting a short transient followed by an exponentially damped oscillation.

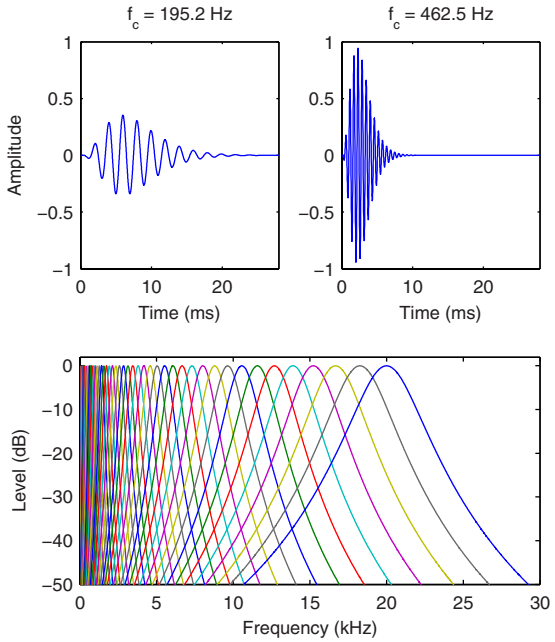


FIG. 1. (Color online) In the upper row the waveforms of two gammatone filters are plotted. The lower row shows the magnitude frequency response of $M = 50$ gammatone filters that are equally distributed along the ERB scale from 20 Hz to 20 kHz.

B. Overcomplete gammatone signal model

To analyze overcomplete gammatone signal models we first have to define a corresponding discrete signal processing system (Fig. 2). The signal $x[n]$ is analyzed with a filterbank where $h_m[n]$, $m \in [0, M-1]$ denotes the impulse responses of M gammatone filters. This splits the full-band signal $x[n]$ into M frequency bands (subbands). In many signal processing applications these subbands are subsampled by decimation factors N_m to remove redundancy from the internal representation and thereby reducing the overcompleteness of the signal model. For the maximally decimated case with $1/N_0 + \cdots + 1/N_{M-1} = 1$, a critical sampling is realized, meaning that the amount of data (samples per second) in the transformed domain and for the original signal is the same. For $\sum_{m=0}^{M-1} 1/N_m > 1$ the signal model is overcomplete, and there are more subband coefficients $y_m[n]$ per time unit than input samples $x[n]$. All subband coefficients $y_m[n]$ are then routed into a subband processing block. In this block, further operations could be performed, for example, a quantization of the subband coefficients controlled by a psychoacoustic model (PAM) or a sparse signal model algorithm like

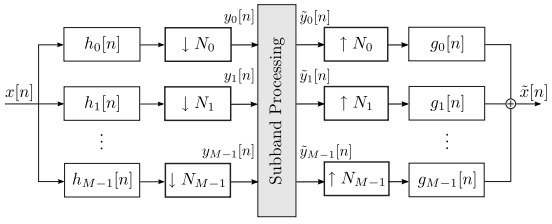


FIG. 2. Discrete signal processing system used to analyze the overcomplete gammatone signal models.

MP (see Appendix A). After the subband processing, the signal $\tilde{x}[n]$ is reconstructed from the M processed subband signals $\tilde{y}_m[n]$ by upsampling with N_m, followed by the synthesis filterbank with the filters having impulse responses $g_m[n]$, $m \in [0, M-1]$.

The analysis presented in this paper is applicable for two different variations in the gammatone signal model. The first variation uses gammatone analysis filters $h_m = \gamma[n]$ and reversed gammatone synthesis filters $g_m = \gamma[-n]$. This is the most commonly used design, for example, in audio coding applications.[13,14,16] The second variation uses reversed gammatone analysis filters $h_m = \gamma[-n]$ and gammatone synthesis filters $g_m = \gamma[n]$. This system can be used to perform a fast MP analysis with a gammatone dictionary (see Appendix A). By choosing the synthesis filters as the time-reverse of the analysis filters the overall filterbank response has a linear phase in both designs.

A gammatone signal model is normally designed to cover only a limited frequency range.[9–14,16,36] Consequently, the analyses in this paper have been conducted using such bandlimited gammatone signal models. We distributed the center frequencies of the gammatone filters equally spaced on the ERB scale within the interval $f_c \in [20, 20\,000]$ Hz, which represents the approximated human hearing range.[37]

III. FRAME-THEORETIC ANALYSIS OF AN OVERCOMPLETE GAMMATONE SIGNAL MODEL

In this section, we will perform a frame-theoretic analysis of the overcomplete gammatone signal model. We will introduce the theory of frames and use it to evaluate the properties of the corresponding frame of a non-decimated and a decimated gammatone signal model. All calculations have been performed with a sampling rate of 96 kHz, and the length of the impulse responses $h_m[n]$ and $g_m[n]$ was 8192 samples or 85.3 ms, respectively.

A. The theory of frames

The theory of frames provides a mathematical framework to analyze overcomplete signal models.[23–25] A *frame* of a vector space $\mathbf{V}$ is a set of vectors $\{\mathbf{e}_m\}$ which satisfy the following *frame condition*:[25]

$$A\|\mathbf{v}\|^2 \leq \sum_m |\langle \mathbf{v}, \mathbf{e}_m \rangle|^2 \leq B\|\mathbf{v}\|^2 \quad \forall\, \mathbf{v} \in \mathbf{V}, \tag{2}$$

with the *frame bounds* $A > 0$ and $B < \infty$. Frames can be seen as a generalization of bases, as the set $\{\mathbf{e}_m\}$ is allowed to be linearly dependent, and Eq. (2) implies that the set $\{\mathbf{e}_m\}$ must span the vector space $\mathbf{V}$. Otherwise it would follow $A = 0$ from $\langle \mathbf{v}, \mathbf{e}_m \rangle = 0$ for $\mathbf{v} \in \mathbf{V} \backslash \mathrm{span}\{\mathbf{e}_m\}$.

The frame condition can also be written as $A\|\mathbf{v}\|^2 \leq \langle \mathbf{S}\mathbf{v}, \mathbf{v} \rangle \leq B\|\mathbf{v}\|^2$ with $\mathbf{S}$ being the *frame operator* defined as

$$\mathbf{S}\mathbf{v} = \sum_m \langle \mathbf{v}, \mathbf{e}_m \rangle \mathbf{e}_m. \tag{3}$$

The frame bound A is the essential infimum and the frame bound B is the essential supremum of the eigenvalues of $\mathbf{S}$.[25] A frame is called *tight* if $B/A = 1$ and *snug* if $B/A \approx 1$. The advantage of a tight frame is that perfect reconstruction can be done by the frame itself:

$$\mathbf{v} = \frac{1}{A} \sum_m \langle \mathbf{v}, \mathbf{e}_m \rangle \mathbf{e}_m \quad \forall\, \mathbf{v} \in \mathbf{V}. \tag{4}$$

The frame bounds for the discrete signal processing system as shown in Fig. 2, are given by the following inequality:

$$A\|\mathbf{x}\|^2 \leq \sum_{m=0}^{M-1} \sum_{k=-\infty}^{\infty} |\langle \mathbf{x}, \mathbf{h}_{m,k} \rangle|^2 \leq B\|\mathbf{x}\|^2 \quad \forall\, \mathbf{x} \in \ell^2(\mathbb{Z}), \tag{5}$$

with $m \in [0, M-1]$, $k \in \mathbb{Z}$, and the vectors $\mathbf{h}_{m,k}$ containing the filter coefficients $h_m(kM-n)$ and $\mathbf{x} \in \ell^2(\mathbb{Z})$ being the vector that contains the input samples $x[n]$.

In general, the smaller the ratio B/A is, the better the numerical properties of the signal model will be. If B/A is close to 1, then the assumption of energy preservation may be used without much error when relating the energy of the subband signals $y_m[n]$ to the energy of the input signal $x[n]$ and the output signal $\tilde{x}[n]$. This is important in audio coding applications, as it guarantees that small quantization errors introduced in the subband signals will result in only small reconstruction errors. It enables a bit allocation optimized for minimum error in the subbands to be near-optimal for the final output signal.

The speed of convergence for algorithms like MP also depends on the frame bounds, as shown in Sec. V. In this context it is to note that the frame realized by a MP decomposition with a dictionary of atoms $\mathbf{e}_k$ is identical to a frame realized by a filterbank with the matched filters $\mathbf{e}_k^*[-n]$, as shown in Appendix A.

The frame operator $\mathbf{S}$ can be represented in the polyphase domain by the $M \times M$ matrix $\mathbf{S}(z) = \tilde{\mathbf{E}}(z)\mathbf{E}(z)$, where $\mathbf{E}(z)$ is the analysis polyphase matrix of the filterbank[38] and the eigenvalues of the frame operator $\mathbf{S}$ equal the eigenvalues $\lambda_n(\theta)$ of the matrix $\mathbf{S}(e^{i\theta}) = \mathbf{E}^H(e^{i\theta})\mathbf{E}(e^{i\theta})$. Bolcskei *et al.*[27] could show that the frame bounds A and B are the essential infimum and essential supremum, respectively, of the eigenvalues $\lambda_n(\theta)$. Thus, the computation of the frame bounds of overcomplete gammatone signal models using their polyphase matrix representations is possible. Note that in the non-decimated case, the frame bounds and respective eigenvalues are related to the ripple in the overall frequency response of the filterbank.

The eigenanalysis of a signal model is only applicable for a limited frequency interval if the corresponding filterbank is non-decimated. For $N_m > 1$, the mapping of the eigenvalues of the frame operator to the analyzed frequency interval is lost. Thereby the essential infimum and essential supremum can only be calculated for the entire frequency range, from zero to half the sampling frequency. This results to a lower frame bound of $A = 0$ for bandlimited signal models, like the here analyzed overcomplete gammatone signal model, where filters do not cover frequencies below 20 Hz and above 20 kHz. To circumvent this problem, we added two additional filters for the frequency intervals not covered by the gammatone filterbank, i.e., a lowpass for the $[0, 20]$ Hz frequency interval and a highpass filter for

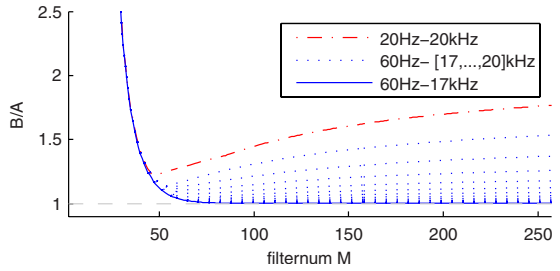


FIG. 3. (Color online) The frame-bound ratios B/A of non-decimated gammatone signal models with the number of filters $M \in [2,256]$ analyzed over the frequency intervals 20 Hz–20 kHz and 60 Hz–[17,20] kHz. For the frequency interval of 60 Hz–17 kHz, the frame-bound ratio converges toward a tight frame for higher filter numbers.

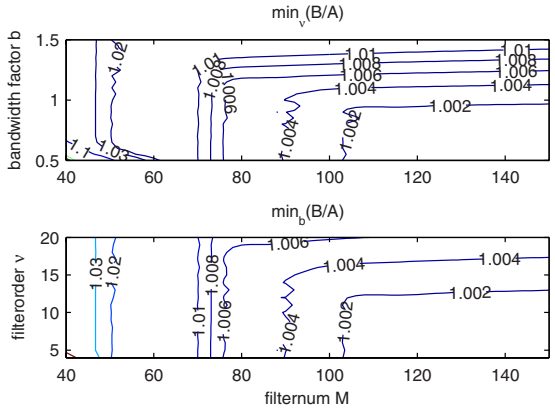


FIG. 4. (Color online) Best possible frame-bound ratios for a fixed bandwidth factor b and filter number M (upper plot) or filter order ν and filter number M (lower plot). The gammatone signal model parameters were $b \in [0.5,1.5]$, $\nu \in [4,20]$, and $M \in [40,150]$ analyzed over the frequency interval from 60 Hz to 17 kHz.

$[20,48]$ kHz. Thereby we could compute A for a decimated gammatone signal model within the limited frequency range. B was computed without additional filters.

B. Analysis of a non-decimated overcomplete gammatone signal model

An overcomplete signal model results in a large quantity of subband coefficients for every filter. To reduce bitcoding and computational costs, it is of interest to know the smallest number M of subbands needed to achieve good frame-bound ratios. As the frame bounds of $\gamma[n]$ are identical to the frame bounds of $\gamma[-n]$, we only need to analyze the frame of the gammatone prototype $\gamma[n]$ itself. The frame bounds A and B of the non-decimated overcomplete gammatone signal model can be computed, as described in Sec. III A, and the respective frame-bound ratios B/A are shown in Fig. 3. The parameters of the analyzed gammatone signal models were $b=1.019$, $\nu=4$ with $M \in [2,256]$ center frequencies between 20 Hz and 20 kHz.

Figure 3 shows that the gammatone signal model does not realize a frame for the frequency interval of its center frequencies. The frame-bound ratio is mainly determined by small eigenvalues of the frame operator $\mathbf{S}$ found at the first and last gammatone filters (see also Fig. 11). The ERB scale distributes the center frequencies of the gammatone atoms in such a way that the overlapping filters result in almost constant eigenvalues. As for the first and the last filters this overlap is not fully realized; the essential infimum of the eigenvalues results in a low lower frame bound A. If we perform the analysis over a reduced frequency interval (see Fig. 3 and Table I), the frame-bound ratio improves and the gammatone signal is able to achieve a snug frame from $M=50$ subbands on. This marginal reduction in the frequency interval is non-critical as it still embeds the class of natural sounds with speech, for example, ranging approximately from 80 Hz to 10 kHz.

For $M=50$ the frame bounds are $A=1.167$ and $B=1.294$, which results in a frame-bound ratio of $B/A=1.109$. This means that, depending on the actual signal, the energy of the input or output signal of the filterbank may be different from the subband energy by a factor between 1.167 and 1.294. For higher filter numbers the frame-bound ratio converges toward a tight frame and for $M=100$ a frame-bound ratio of $B/A=1.003$ is achieved.

For applications that allow a deviation from the human gammatone parameters, we also analyzed the influence of the bandwidth parameters $b \in [0.5,1.5]$ and the filter orders $\nu \in [4,20]$ on the frame-bound ratio for the frequency interval from 60 Hz to 17 kHz (see Fig. 4). For $M=50$ gammatone atoms, the best frame-bound ratio $B/A=1.020$ is achieved for a filter order $\nu=11$ and the bandwidth factor $b=0.85$. For $M=100$ the filter order $\nu=12$ and the bandwidth factor $b=0.5$ result in the lowest frame-bound ratio of $B/A=1.003$. The contour plot in Fig. 4 shows that these best frame-bound ratios are located in relatively shallow minima. More generally, we can conclude that for a filter number of $M=50$, snug frames can be achieved with $b>0.7$ and all examined filter orders. For $M=100$ a tight frame is possible with $b \leq 1$, $\nu<13$. Additionally it can be seen that for a small number of filters ($M<50$) larger bandwidths achieve better frame-bound ratios. More interestingly, for a higher number of filters, large filter bandwidths introduce a decline in the frame-bound ratio which is explained in detail in Sec. VI and Fig. 11.

TABLE I. Frame-bound ratios B/A analyzed for different bandlimited signals and number of gammatone filters M.

M	Frequency interval	B	A	B/A	Frame
50	[20 Hz, 20 kHz]	1.294	1.046	1.238	Not snug
50	[40 Hz, 17 kHz]	1.294	1.167	1.109	Snug
100	[20 Hz, 20 kHz]	2.462	1.697	1.451	Not snug
100	[60 Hz, 17 kHz]	2.462	2.455	1.003	≈ tight

C. Analysis of a decimated overcomplete gammatone signal model

To further reduce encoding and subband processing costs, it is often favorable to remove the redundancy in an overcomplete signal model by downsampling its subband coefficients by factors $N_m>1$. The decimation of the filterbank coefficients can result in distortions, which will worsen the frame-bound ratio of the decimated signal model. Thus, a

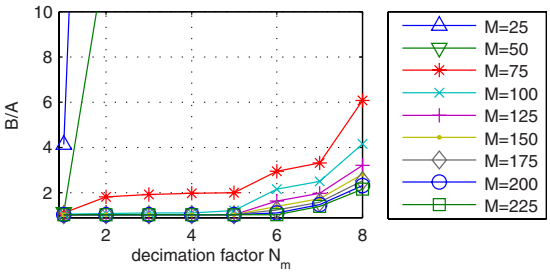


FIG. 5. (Color online) The frame-bound ratios B/A of decimated gammatone signal models with the number of filters $M \in \{25,50,\ldots,200,225\}$ and decimation factors $N_m \in [1,8]$ analyzed over the frequency interval of 60 Hz–17 kHz.

frame-theoretic analysis can be used to analyze the introduced distortions for different decimation factors N_m. We derived frame bounds for a decimated overcomplete gammatone signal model for the frequency interval from 60 Hz to 17 kHz by introducing additional filters to allow the derivation of A, as described in Sec. III A. The resulting frame-bound ratios B/A are shown in Fig. 5. It can be seen that no snug frame can be achieved for $M \leq 75$ filters with an equal decimation of the subband coefficients. For higher filter numbers, a snug frame can be realized up to an equal decimation of the subband coefficients of $N_m=4$, $N_m=5$, and $N_m=6$ for the filter numbers $M=100$, $M \in [125,150]$, and $M \in [175,255]$, respectively.

To derive optimal decimation factors for an overcomplete gammatone signal model, a full search over all possible N_m by computing the corresponding frame-bound ratios would be necessary, which is computationally intractable. It is further to note that distortions that fall into a frequency range where the signal has only little energy will have a minor effect compared to distortions in frequency bands, where most of the signal energy is present. This cannot be exploited by an optimization based on frame-bound ratios due to the lost mapping of the eigenvalues of the frame operator to the analyzed frequency interval. Therefore we introduce and use in Sec. IV an alternative technique to derive optimal decimation factors.

IV. BIFREQUENCY ANALYSIS OF A DECIMATED OVERCOMPLETE GAMMATONE SIGNAL MODEL

To allow the optimization of decimation factors dependent on the applied signal, we will introduce in this section the bifrequency analysis[39] and define a SAR. The bifrequency analysis has the additional advantage that it offers a complete frequency description of the distortions introduced by a decimation of the subband coefficients. This leads to a better insight of the design limitations, i.e., to Conditions I and II as given below. This allows to reduce the computational costs of an optimization of the decimation factors. All results in this section were derived with a sampling rate of 44.1 kHz, which is a common sampling rate in signal processing applications like audio coding. The length of the analyzed impulse responses $h_m[n]$ and $g_m[n]$ has been set to 4096 samples or 92.9 ms, respectively.

A. Bifrequency analysis

An alternative theoretical analysis of the decimated gammatone signal models is possible by the fact that a decimated filterbank can also be understood as a linear time-varying (LTV) system

$$\mathbf{y}[n_y] = \sum_{n_x=-\infty}^{\infty} \mathbf{k}[n_y,n_x]\mathbf{x}[n_x], \tag{6}$$

with a periodic system response $\mathbf{k}[n_y,n_x]=\mathbf{k}[n_y+\ell N, n_x+\ell N]$, $\ell \in \mathbb{Z}$, where $\mathbf{x}[n_x]$ is the input and $\mathbf{y}[n_y]$ is the output sequence. $\mathbf{k}[n_y,n_x]$ denotes the response of the system at the discrete time n_y to a unit sample applied at discrete time n_x. For periodic LTV systems, a bifrequency analysis[39] gives a complete description of the system as well as of its aliasing components. The discrete bifrequency system function[26] is defined as

$$\mathbf{K}[e^{i\omega_y},e^{i\omega_x}] := \frac{1}{2\pi}\sum_{n_y=-\infty}^{\infty}\sum_{n_x=-\infty}^{\infty}\mathbf{k}[n_y,n_x]e^{i\omega_x n_x}e^{-i\omega_y n_y}, \tag{7}$$

relating the input signal spectrum $\mathbf{X}[e^{i\omega_x}]$ to the output signal spectrum $\mathbf{Y}[e^{i\omega_y}]$ with

$$\mathbf{Y}[e^{i\omega_y}] = \int_{-\pi}^{\pi}\mathbf{K}[e^{i\omega_y},e^{i\omega_x}]\mathbf{X}[e^{i\omega_x}]d\omega_x. \tag{8}$$

In the analyzed gammatone signal models, the only periodically time-varying parts are the decimators and interpolators. Therefore, the overall bifrequency map is composed of nonzero unity-slope parallel lines with a constant factor, on whose input and output spectra the effects of the analysis and the synthesis filters, respectively, are projected.[40] The center line represents the time-invariant part of the system; all other lines represent the parts of the system which cause aliasing (see also Fig. 6). As an objective measure of the aliasing distortions in a signal model we used a signal-to-alias (SAR), defined analogous to the commonly used signal-to-noise ratio (SNR). For a given input signal spectrum $\mathbf{X}[e^{i\omega_x}]$ the SAR is defined as

$$\mathrm{SAR}(\mathbf{X}[e^{i\omega_x}]) = -10\log_{10}\left(\frac{T_1^2}{\sum_{n\in\{N_m\}}T_n^2}\right), \tag{9}$$

with

$$T_n = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\delta(n\omega_x-\omega_y)\mathbf{K}[e^{i\omega_y},e^{i\omega_x}]\mathbf{X}[e^{i\omega_x}]d\omega_x d\omega_y, \tag{10}$$

and $\delta(\cdot)$ being the Dirac pulse. The time-invariant part of the system corresponds to T_1, and the aliasing components of the LTV system are represented by the T_n.

To avoid in-band aliasing distortions, N_m must be chosen in such a way that all integer multiples of the decimated Nyquist frequency lie outside the mth passband of a subband [see Fig. 6(b)]. For an aliasing-free signal model this results in the following necessary condition to prevent in-band aliasing.

Condition I. With ω_m^L and ω_m^H being the starting and stopping cutoff frequencies of the mth gammatone filter ($0 \leq \omega_m^L \leq \omega_m^H \leq \pi$) it needs to hold

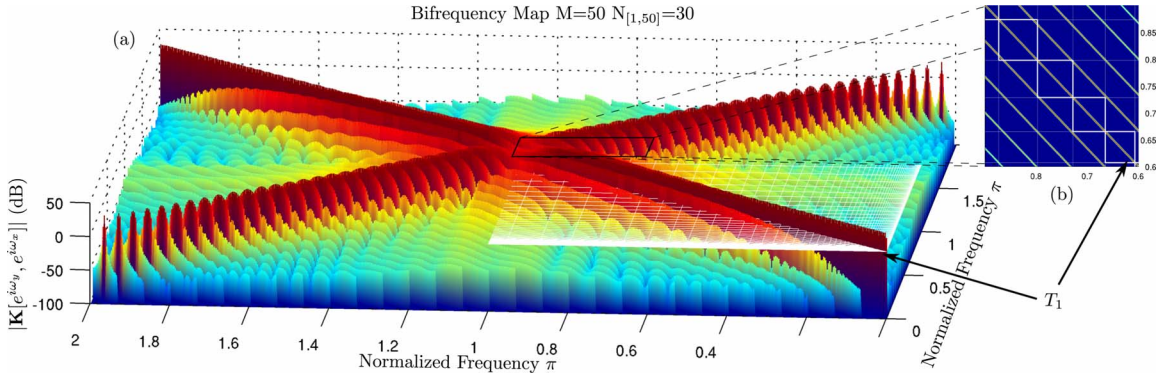


FIG. 6. (Color online) (a) Bifrequency map for a gammatone signal model with the number of filters $M=50$ and the decimation factor of $N_m=30$ in every subband. The axes show the normalized frequency domains associated with the input and output signals. The center line represents the time-invariant part (T_1) that maps the input to the output signal and is independent of any decimation. All other lines are due to aliasing terms ($T_{n>1}$) introduced by a decimation of the subband coefficients. The zoom-in (b) shows that in this example in-band aliasing occurs in the last three filters, in which aliasing components fall into the passband of these filters. The filter's passbands are indicated by a grid of thin white lines.

$$(k\pi/N_m) \notin [\omega_m^L, \omega_m^H] \quad \forall\, k \in \mathbb{N}. \tag{11}$$

This dependency on the bandwidth of the corresponding gammatone filter limits the possible decimation factors to the set which fulfills $N_m < \pi/(\omega_m^H - \omega_m^L)$. In contrast to an ideal bandpass filter, which has a discontinuity in magnitude at the cutoff frequencies, real filters like the gammatone filter exhibit a magnitude response that changes gradually from the passband to the stopbands. A commonly chosen decrease in magnitude to define the cutoff frequency is an attenuation of 3 dB.

Inter-band aliasing can be reduced if the decimation factors are chosen in such a way that an aliasing term of a filter in one subband can be canceled by another aliasing term of a filter in another subband. Such a set of integer decimation factors N_m in which each aliasing term occurs at least twice is called a *compatible set*[38,41,42] and needs to fulfill the following condition.

Condition II. Let $L := \mathrm{lcm}(\{N_m\}_{m=0}^{M-1})$ be the least common multiplier (lcm) of the set of decimation factors $\{N_m\}_{m=0}^{M-1}$. If the set is an apposition of repeated distinct integers $\{\mathcal{N}_1, \mathcal{N}_1, \ldots, \mathcal{N}_1, \ldots, \mathcal{N}_{K-1}, \ldots, \mathcal{N}_{K-1}\}$ with $\mathcal{N}_j \in \{N_m\}_{m=0}^{M-1}$ and n_j denoting the number of $\mathcal{N}_j$ in this set, then it needs to hold

$$\min\left\{ \frac{\mathrm{lcm}\left(\frac{L}{\mathcal{N}_i}, \frac{L}{\mathcal{N}_j}\right)}{\frac{L}{\mathcal{N}_j}} \right\}_{\substack{i=0 \\ i\neq j}}^{M-1} - 1 < n_j. \tag{12}$$

B. Analysis of a decimated overcomplete gammatone signal model

We will use the results from Sec. IV A to show how optimal decimation factors N_m for a given decimated overcomplete gammatone signal model and a given signal spectrum $\mathbf{X}[e^{i\omega_x}]$ can be derived. Let $\mathbf{N} := (N_0, N_1, \ldots, N_{M-1}) \in [1, M-1]^M$ be the M-dimensional vector space of all possible decimation factors for a gammatone signal model. We can reduce the size of $\mathbf{N}$ by allowing only decimation factors that fulfill Conditions I and II. The cutoff frequency was set at 3 dB stopband attenuation. The size of the set of possible decimation factors can be further reduced using the constraint $N_0 \geq N_1 \geq \cdots \geq N_{M-1}$, which is derived from Condition I and the fact that the gammatone signal model has monotone increasing bandwidths. To select decimation factors that form a *compatible set*, the decimation factors can be required to be powers of 2.

To derive for a given degree of overcompleteness $O = \sum_{m=0}^{M-1} 1/N_m$, a set of decimation factors with minimal aliasing distortions, the SAR can be used as a quality measure. To exemplify this, we analyzed an overcomplete gammatone signal model with $M=50$ filters, center frequencies ranging from 20 Hz to 20 kHz, and $N_m \in \{1, 2, 4, 8, 16, 32, 64, 128, 256, 512\}$. We further evaluated if varying the bandwidth of the gammatone filters has an influence on the aliasing distortions. Analyzing Fig. 6, it can be seen that the major aliasing distortions occur in the high-frequency bands due to the nonuniform frequency resolution of the gammatone signal model. For applications like speech or audio coding, where only a small amount of signal energy falls in the high-frequency bands, these distortions will have a minor effect compared to the distortions in the low-frequency band, where most of the signal energy is present. Therefore it is favorable to optimize the decimation factors according to the SAR computed for the specific spectrum of the applied signal class. In this example we used the spectrum of the audio test signal "Tom's Diner" by Vega (svega.wav). Table II shows the SAR achieved by optimal decimation factors (stated in Appendix B), selected from a set of decimation factors that is constructed as described above and that results in the degrees of overcompleteness $O = 1, 2, \ldots, 8$, respectively. They are compared with commonly chosen decimation factors that are inverse-proportional to the bandwidth of the gammatone filters while fulfilling Condition I. The optimized decimation factors achieve a SAR improvement of 4.7 dB on average compared to the commonly chosen decimation factors. This can be seen as a significant improvement, recalling that a SAR improvement of 6 dB means a reduction in the distortion energy due to aliasing components by a factor of 2. As the overcomplete gammatone signal model realizes for $M=50$ only a snug frame, we additionally investigated if the SAR can be improved using different filter

TABLE II. The SAR for svega.wav and a gammatone signal model with $M=50$ filters achieved with optimized decimation factors compared to commonly chosen decimation factors that are inverse-proportional to the bandwidth of the filters while fulfilling Condition I.

SAR (dB)	$O=1$	$O=2$	$O=3$	$O=4$	$O=5$	$O=6$	$O=7$	$O=8$
Optimized N_m	9.5	14.2	15.6	17.5	18.2	18.5	18.9	19.2
Prop. bandwidth	6.2	8.1	10.6	11.3	13.4	14.5	14.6	15.7

bandwidths. It showed that for $M=50$ a deviation from the human bandwidth parameter $b=1.019$ can reduce inter-band aliasing distortions from 1 up to 15.2 dB for $O=1$ and $O=8$, respectively (see Fig. 7). As an increase in the filter bandwidth leads to an increase in the energy in the aliasing components, this reduction in aliasing distortions can be addressed to an optimized cancellation of aliasing terms. So depending on the number of applied gammatone filters, the bandwidth factor b should also be included into the optimization process.

V. APPLICATIONS

In this section we report on the signal reconstruction performance of overcomplete gammatone signal models using the example of audio coding and compare the findings with the theoretical results from Secs. III and IV. We applied a coding scheme whose block diagram is shown in Fig. 2.

In the first experiment, we investigated the signal reconstruction and subband algorithm performance of a nondecimated overcomplete gammatone signal model ($N_m=1$), as analyzed in Sec. III. We tested two signal model variations. In the first variation (GTFB), we evaluated the standard overcomplete gammatone signal model with $h_m=\gamma[n]$, $g_m=\gamma[-n]$ and without subband processing. In the second variation, a sparse overcomplete gammatone signal model was realized with $h_m=\gamma[-n]$, $g_m=\gamma[n]$, and a MP algorithm[19] was performed in the subband processing block. The stopping condition was set to 2000 atoms/s and it was implemented as described in Appendix A. The test signal for this initial audio coding experiment was the commonly used Tom's Diner by Suzanne Vega (svega.wav). In accordance with the theoretically derived results (Fig. 3), the signal reconstruction error decreased for both schemes with an increasing number of filters and saturated for higher filter numbers (Fig. 8). For the overcomplete gammatone signal model (GTFB), near-perfect reconstruction was achieved for $M\geq 100$. For the sparse overcomplete gammatone signal model (MP) the SNR rose to 22.5 dB at $M\approx 70$ and continued to slightly improve further for higher filter numbers until it stayed constant at 23.5 dB for $M\geq 500$ gammatone filters. This shows that the convergence speed of the MP algorithm facilitated also from small frame-bound ratio improvements close to $B/A=1$, as the overcomplete gammatone signal model did not contribute further to the signal reconstruction for $M\geq 100$.

We further evaluated a basic perceptual audio coding scheme by scaling the subband coefficients $y_m[n]$ according to a PAM before performing a fixed quantization.[43] The PAM was realized by the MPEG-2 AAC/MPEG-4 audio standard reference implementation,[44] and a linear 7 bit quantizer was used. The coding and decoding of the scaled and quantized coefficients were assumed to be lossless and therefore omitted. Finally according dequantization and rescaling was performed before the audio signal was reconstructed using the synthesis filterbank. We measured the perceived audio quality of the resulting audio signals relative to the original test signal using a model of auditory perception (PEMO-Q).[45] The estimated perceived audio quality was mapped to a single quality indicator, the objective difference grade (ODG).[46] This is a continuous scale from 0 for "imperceptible impairment," −1 for "perceptible but not annoying impairment," −2 for "slightly annoying impairment," −3 for "annoying impairment" to −4 for "very annoying impairment." As explained in Sec. III, subband processing algorithms like perceptual audio coding rely on the assumption of energy preservation in the signal model. Their performance therefore depends on the achieved frame-bound ratio of the used signal model. As shown in Fig. 9, the GTFB signal model without quantization achieved transparent audio coding from $M>55$ gammatone filters on. Linearly quantizing the subband coefficients to a 7 bit encoding, the ODG converged around $M>45$ to approximately −2.5. Scaling the important subband coefficients before quantization according

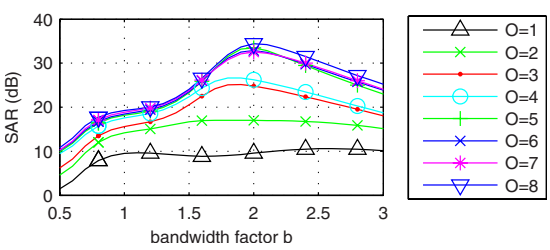


FIG. 7. (Color online) The SAR achieved by optimized decimation factors for a given degree of overcompleteness O and different bandwidth factors b of $M=50$ gammatone filters.

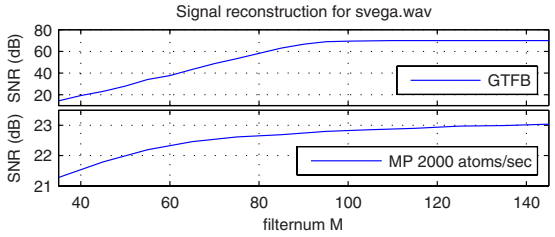


FIG. 8. (Color online) Signal reconstruction experiment using nondecimated overcomplete gammatone signal models for the svega.wav test signal. The upper plot shows the results for a signal model without subband processing (GTFB) and the lower plot shows the achieved SNR for a sparse gammatone signal model based on the MP algorithm.

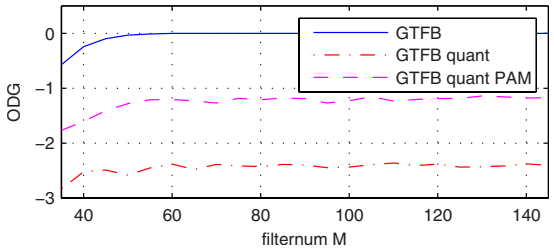


FIG. 9. (Color online) Perceptual reconstruction quality for svega.wav encoded without quantization with a linear quantization and a linear quantization including a PAM.

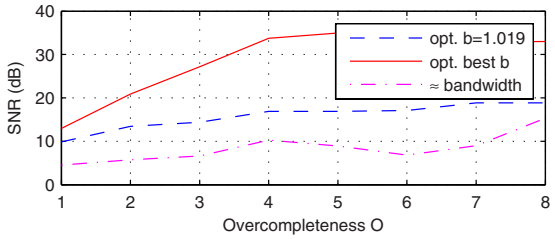


FIG. 10. (Color online) Signal reconstruction experiment using decimated overcomplete gammatone signal models being optimized to maximize the SAR of the test signal (svega.wav), as described in Sec. IV B, compared to commonly chosen decimation factors that are inverse-proportional to the bandwidth of the filters while fulfilling Condition I.

to a PAM showed an improvement in the perceived audio quality until $M \approx 60$ where an ODG of approximately -1.2 is achieved. With the results from Sec. III it can be concluded that for audio coding applications at least a snug frame should be realized by the gammatone signal model. Clearly, to further improve the quality up to an ODG of zero, finer quantization is needed.

In the second experiment, we investigated the signal reconstruction performance of decimated overcomplete signal models with $M=50$ filters and without any subband processing. As a reference signal model we selected commonly chosen decimation factors that are inverse-proportional to the bandwidth of the gammatone filters, while fulfilling Condition I. We compared their achieved signal reconstruction performance with optimized decimation factors for a gammatone signal model having a fixed bandwidth factor $b=1.019$ and for a gammatone signal model where also the bandwidth of the filters was optimized, as described in Sec. IV B. The audio test file was svega.wav, and the results are plotted in Fig. 10. It can be seen that the decimation factors optimized to maximize the SAR of the audio signal as described in Sec. IV B result in a better SNR than the N_m that are increased proportional to the filter bandwidth and fulfill Condition I. It further shows that for the snug frame realized with $M=50$ gammatone filters, a deviation from the human bandwidth parameter $b=1.019$, if allowed in the context of the application, can reduce the aliasing distortions and improve the signal reconstruction performance.

VI. DISCUSSION

Applications that use an overcomplete gammatone signal model can be divided into two groups. The first group is concerned with modeling the auditory system. In these studies, the number of auditory filters is inferred from a reasonable filter spacing determined by the estimated bandwidths of the auditory filters. A common value used is 1 filter per ERB,[9,10,45] which results in 39 filters for the human cochlea, whose basal end corresponds to 38.9 on the ERB scale.[47] The second group of applications is concerned with signal processing tasks, for example, audio coding and speech recognition. Hereby, not an accurate replication of the auditory system is strictly needed, but a maximal performance of the algorithm is desired. Therefore, the number of gammatone filters should be chosen optimizing the performance of the subsequent processing algorithms and the introduced computational load. Most signal processing applications using an overcomplete gammatone signal model so far have used psychoacoustically derived filter numbers, which do not result in a frame (see Table III). As shown in Sec. V, subband processing algorithms like MP or a perceptual quantizer show an improved performance for improved frame bounds.

Note that it is not self-evident that an overcomplete gammatone signal model can achieve a snug frame and converge to a tight frame. The parameters of the gammatone function have been derived from psychoacoustic experiments and are not specifically designed to realize a frame in the mathematical sense. Further analysis of the eigenvalues showed that at higher filter numbers ($M>60$), the frame-bound ratio is determined mainly by the fact that the frequency spacing of the ERB scale does not fully match the filter overlap to the filter bandwidths. This introduces a positive shift of the largest eigenvalues toward higher frequencies (see Fig. 11). Therefore we evaluated if marginal alter-

TABLE III. Examples for gammatone signal model parameters found in the literature. The frame-bound analysis was performed on a limited frequency interval to exclude distortion effects from the first and last filters.

Paper	Interval of center frequencies	M	Given rational	Filter per ERB	B/A	Frame-bound analysis interval	Frame
Ambikairajah *et al.*[36]	50 Hz–7.0 kHz	21	"Ripple within 1.5 dB"	⋯	1.481	100 Hz–7 kHz	Not snug
Brucke *et al.*[56]	73 Hz–6.7 kHz	30	1 filter per ERB	1.0	1.322	70 Hz–6.2 kHz	Not snug
Feldbauer *et al.*[16]	100 Hz–3.6 kHz	50	Frame-bound ratio	2.2	1.003	150 Hz–3.0 kHz	≈ tight
Hohmann[17]	70 Hz–6.7 kHz	30	1 filter per ERB	1.0	1.332	65 Hz–6.3 kHz	Not snug
Kubin and Kleijn[14]	100 Hz–3.6 kHz	20	"Physiologically-motivated"	0.9	1.364	190 Hz–3.1 kHz	Not snug
Lin *et al.*[15]	<4 kHz	25	Not stated	0.9	1.572	35 Hz–4.0 kHz	Not snug
Ma *et al.*[57]	50 Hz–8.0 kHz	64	"Computational costs"	2.0	1.003	100 Hz–6.2 kHz	≈ tight
This study	20 Hz–20.0 kHz	50	Frame-bound ratio	1.2	1.109	60 Hz–17 kHz	Snug
		100	Frame-bound ratio	2.4	1.003	60 Hz–17 kHz	≈ tight

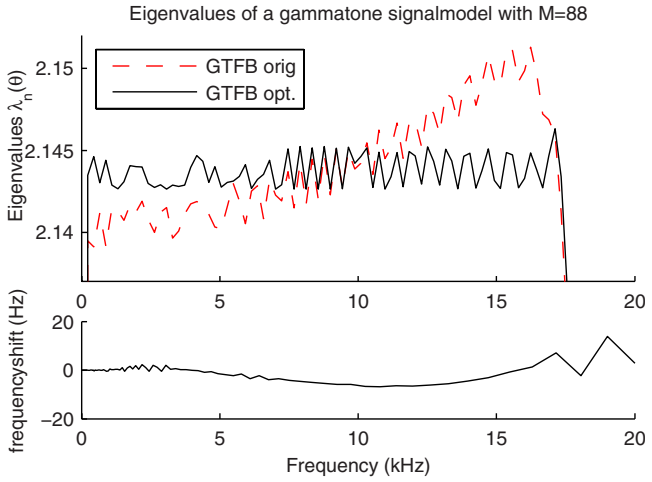


FIG. 11. (Color online) The eigenvalues $\lambda_n(\theta)$ of an overcomplete gammatone signal model with $M=88$ filters being equally spaced on the ERB scale compared to an optimized frequency scale with frequency shifts applied to the ERB scale as shown in the lower row.

ations of the filter's center frequency can improve the gammatone signal model. Using the frame-bound ratio as a cost function, a standard optimization algorithm like the MATLAB function fmincon can be used to derive the frequency shifts necessary to remove the monotonic shift. The derived frequency shifts reduced the center frequencies slightly at middle frequencies, compensating this with a frequency increase at the lower and higher frequencies, see also the example shown in Fig. 11. For this example with $M=88$ the frame-bound ratio could be improved from 1.006 to 1.001 by applying only, relative to the center frequency, marginal frequency shifts. Note that these results are only of theoretical interest, as the gammatone signal already forms an almost tight frame at higher filter numbers M, and the derived optimization does not improve the numerical properties of the signal model at a noticeable level. So for the overcomplete gammatone signal model, the ERB scale itself is already close to the frequency tiling of the time-frequency plane that achieves the best frame-bound ratio.

For a decimated overcomplete gammatone signal model, the derived frame bounds cannot be used to optimize the decimation factors in dependency of the signal spectrum, as explained in Sec. IV. Another possibility to evaluate such bandlimited signal models is the computation of the SAR allowing the optimization of the trade-off between linear amplitude distortions and the amount of aliasing. We could show that the common approach to use decimation factors that are proportional to the bandwidth of the filters is suboptimal. The SAR can easily be computed using a two–dimensional fast Fourier transform (2D-FFT), and we therefore recommend for signal processing applications using a decimated overcomplete gammatone signal model to utilize decimation factors N_m being optimized for the applied signal class.

Note that very long finite-impulse responses and high sampling rates have been used in this study to derive frame bounds that are valid approximations for the analog gammatone filters. Applications using other digital realizations of the gammatone filterbank like infinite-impulse response filters might result in slightly different frame bounds.[48]

A linear gammatone signal model is a valid approximation of the human auditory filters for moderate sound pressure levels. It has been shown that the filter shape of the auditory filter changes with stimulus level,[49] which led to the development of dynamic, non-linear auditory filter models.[50,51] The analysis methods applied in this study cannot directly be applied to such dynamic filters and are therefore not within the scope of this manuscript.

VII. CONCLUSIONS

Using the theory of frames we could derive that from 2.4 filters per ERB on, a non-decimated overcomplete gammatone signal model achieves near-perfect signal reconstruction and that from $M=55$ (1.3 filters per ERB) filters on, a perceptual transparent audio coding is possible. We further showed that by computing a SAR, the decimation factors in multi-rate signal processing schemes can be optimized, balancing the amplitude and aliasing distortions. We showed for an audio test signal that hereby significant improvements can be achieved.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their constructive comments and corrections, which significantly improved the quality of this manuscript. This work was partly funded by the German Science Foundation (DFG) through the International Graduate School for Neurosensory Science and Systems and the SFB/TRR 31: "The Active Auditory System." The author Stefan Strahl wants to especially thank Astrid Klinge for the inspiring scientific discussions about the manuscript.

APPENDIX A: MP WITH MATCHED FILTERS

MP (Ref. 19) assumes an additive signal model of the form

$$\mathbf{x}=\sum_{i=1}^{K} s_i\mathbf{a}_i, \tag{A1}$$

with the signal vector $\mathbf{x}\in\mathbb{R}^{N\times 1}$, the coefficients $\mathbf{s}=(s_1,s_2,\ldots,s_K)\in\mathbb{C}^K$, and the atoms $\mathbf{A}=(\mathbf{a}_1,\mathbf{a}_2,\ldots,\mathbf{a}_M)\in\mathbb{C}^{N\times M}$ having unit-norm. For an overcomplete signal model, the MP algorithm searches for the sparsest encoding in the infinite number of possible encodings. As mentioned in the Introduction, this sparse signal model resembles the signal analysis performed by the human cochlea.

The algorithm performs a greedy iterative search by selecting at the ith iteration the atom having the largest inner product with the residual $\mathbf{r}_i$:

$$s_{m_i}=\arg\max_{\mathbf{a}_{m_i}\in\mathbf{A}}|\langle\mathbf{r}_i,\mathbf{a}_{m_i}\rangle|^2, \tag{A2}$$

with m_i being the dictionary index of the selected atom at the ith iteration. The new residual is then computed with

$$\mathbf{r}_{i+1}=\mathbf{r}_i-s_{m_i}\mathbf{a}_{m_i}. \tag{A3}$$

If we rewrite the inner products in Eq. (A2) as

$$s_m = \langle \mathbf{r}_i, \mathbf{a}_m \rangle = \sum_{n=1}^{N} r_i[n] \cdot a_m[n] = \sum_{n=1}^{N} r_i[n] \cdot \tilde{a}_m[N-n+1]$$

$$\text{with } \tilde{a}_m[n] = a_m^*[-n] = r_i[n] * \tilde{a}_m[n],$$

it can be seen that the inner products can also be computed using the time reversed atom $\tilde{a}_m$, which is also called a *matched filter*. So we can efficiently compute all inner products using a time-reversed gammatone filterbank. In practical applications of MP the support L of the atoms is often much smaller than the length N of the signal. Therefore most implementations[52–54] divide the signal into overlapping blocks of length L and stepwidth S. With this iterative procedure, only the correlations of the $2L/S-1$ signal blocks which have been altered in the previous iteration need to be recomputed. Using the matched-filter approach we can compute the new correlations of the $2L/S-1$ signal blocks in one step by convolving the $2L$ samples of the whole block once with the matched filterbank. So for a signal of length N and a dictionary size M, we can perform the MP iteration in $\mathcal{O}(MN)$. If MP is performed with a pure gammatone dictionary, we can accelerate the MP algorithm further by precomputing the representations of the gammatone atoms in the filterbank domain and performing the update of the inner products by a simple subtraction in the filterbank domain. For a dictionary of size M, instead of $6M \cdot 2L$ multiplication and $10M \cdot 2L$ additions,[55] the update of the correlations can be done with $M2L$ subtractions.

APPENDIX B: OPTIMAL DECIMATION FACTORS

In Sec. IV B derived optimal decimation factors for svega.wav, $b=1.019$, and $M=50$ are as follows:

$O=1$ $N_{1-10}=128$, $N_{11-33}=64$, $N_{34-49}=32$, $N_{50}=16$,

$O=2$ $N_{1-8}=64$, $N_{9-36}=32$, $N_{37-48}=16$, $N_{49-50}=8$,

$O=3$ $N_{1-24}=32$, $N_{25-40}=16$, $N_{41-50}=8$,

$O=4$ $N_{1-10}=32$, $N_{11-31}=16$, $N_{32-50}=8$,

$O=5$ $N_{1-2}=32$, $N_{3-31}=16$, $N_{32-44}=8$, $N_{45-50}=4$,

$O=6$ $N_{1-20}=16$, $N_{21-44}=8$, $N_{45-49}=4$, $N_{50}=2$,

$O=7$ $N_{1-14}=16$, $N_{15-39}=8$, $N_{40-49}=4$, $N_{50}=2$,

$O=8$ $N_{1-10}=16$, $N_{11-39}=8$, $N_{40-46}=4$, $N_{47-50}=2$.

[1] J. B. J. Fourier, *Théorie Analytique de la Chaleur (The Analytical Theory of Heat)* (Didot, Paris, 1822).
[2] D. Gabor, "Theory of communications," J. Inst. Electr. Eng. **93**, 429–457 (1946).
[3] J. Morlet, G. Arens, I. Fourgeau, and D. Giard, "Wave propagation and sampling theory," Geophysics **47**, 203–236 (1982).
[4] M. Lewicki, "Efficient coding of natural sounds," Nat. Neurosci. **5**, 356–363 (2002).
[5] E. Smith and M. Lewicki, "Efficient auditory coding," Nature (London) **439**, 978–982 (2006).
[6] S. Strahl and A. Mertins, "Sparse gammatone signal model optimized for English speech does not match the human auditory filters," Brain Res. **1220**, 224–233 (2008).
[7] R. Patterson and B. Moore, "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency Selectivity in Hearing*, edited by B. Moore (Academic, London, 1986), pp. 123–177.
[8] R. Patterson, I. Nimmo-Smith, J. Holdsworth, and P. Rice, "An efficient auditory filterbank based on the gammatone function," Paper presented at a meeting of the IOC Speech Group on Auditory Modelling at RSRE, December 14–15, 1987.
[9] T. Dau, D. Püschel, and A. Kohlrausch, "A quantitative model of the effective signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622 (1996).
[10] T. Dau, D. Püschel, and A. Kohlrausch, "A quantitative model of the effective signal processing in the auditory system. II. Simulations and measurements," J. Acoust. Soc. Am. **99**, 3623–3631 (1996).
[11] R. Patterson, "Auditory images: How complex sounds are represented in the auditory system," Acoust. Sci. & Tech. **21**, 183–190 (2000).
[12] M. Cooke, "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573 (2006).
[13] G. Kubin and W. Kleijn, "Multiple-description coding (MDC) of speech with an invertibleauditory model," in Proceedings of the IEEE Workshop on Speech Coding (1999), pp. 81–83.
[14] G. Kubin and W. Kleijn, "On speech coding in a perceptual domain," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (1999), pp. 205–208.
[15] L. Lin, W. Holmes, and E. Ambikairajah, "Auditory filter bank inversion," in Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS) (2001), Vol. **2**, pp. 537–540.
[16] C. Feldbauer, G. Kubin, and W. Kleijn, "Anthropomorphic coding of speech and audio: A model inversion approach," EURASIP J. Appl. Signal Process. **9**, 1334–1349 (2005).
[17] V. Hohmann, "Frequency analysis and synthesis using a gammatone filterbank," Acta. Acust. Acust. **88**, 433–442 (2002).
[18] T. Irino and M. Unoki, "A time-varying, analysis/synthesis auditory filterbank using the gammachirp," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (1998), Vol. **6**, pp. 3653–3656.
[19] S. Mallat and Z. Zhang, "Matching pursuit in a time-frequency dictionary," IEEE Trans. Signal Process. **41**, 3397–3415 (1993).
[20] Z. Cvetkovic and M. Vetterli, "Overcomplete expansions and robustness," in Proceedings of the IEEE International Symposium on Time-Frequency and Time-Scale Analysis (1996), pp. 325–328.
[21] H. Bolcskei and F. Hlawatsch, "Oversampled filter banks: Optimal noise shaping, design freedom, and noise analysis," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (1997), Vol. **3**, pp. 2453–2456.
[22] R. Duffin and A. Schaeffer, "A class of nonharmonic Fourier series," Trans. Am. Math. Soc. **72**, 341–366 (1952).
[23] I. Daubechies, A. Grossmann, and Y. Meyer, "Painless nonorthogonal expansions," J. Math. Phys. **27**, 1271–1283 (1986).
[24] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," IEEE Trans. Inf. Theory **36**, 961–1005 (1990).
[25] I. Daubechies, *Ten Lectures on Wavelets* (SIAM, Philadelphia, PA, 1992).
[26] R. Crochiere and L. Rabiner, *Multirate Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1983).
[27] H. Bolcskei, F. Hlawatsch, and H. Feichtinger, "Frame-theoretic analysis of oversampled filter banks," IEEE Trans. Signal Process. **46**, 3256–3268 (1998).
[28] J. Flanagan, "Models for approximating basilar membrane displacement," J. Acoust. Soc. Am. **32**, 937 (1960).
[29] P. I. Johannesma, "The pre-response stimulus ensemble of neurons in the cochlear nucleus," in *Symposium on Hearing Theory* (Institute for Perception Research, Eindhoven, Holland, 1972), pp. 58–69.
[30] E. de Boer, "On the principle of specific coding," ASME J. Dyn. Syst., Meas., Control **95**, 265–273 (1973).
[31] A. M. H. J. Aertsen and P. I. M. Johannesma, "Spectro-temporal receptive fields of auditory neurons in the grassfrog," Biol. Cybern. **38**, 223–234 (1980).
[32] T. Irino, "An optimal auditory filter," in Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) (1995), pp. 198–201.
[33] B. Moore, R. Peters, and B. Glasberg, "Auditory filter shapes at low center frequencies," J. Acoust. Soc. Am. **88**, 132–140 (1990).
[34] B. Moore and B. Glasberg, "A revision of Zwicker's loudness model,"

Acta. Acust. Acust. **82**, 335–345 (1996).
[35]A. Bell and T. Sejnowski, "Learning the higher order structure of a natural sound," Network Comput. Neural Syst. **7**, 261–266 (1996).
[36]E. Ambikairajah, J. Epps, and L. Lin, "Wideband speech and audio coding using gammatone filter banks," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2001), pp. 773–776.
[37]ISO, ISO 389-7, Acoustics-reference zero for the calibration of audiometric equipment—Part 7: Reference threshold of hearing under free-field and diffuse-field listening conditions, International Organization for Standardization, Geneva (1996).
[38]P. Vaidyanathan, *Multirate Systems and Filter Banks* (Prentice-Hall, Upper Saddle River, NJ, 1993).
[39]L. Zadeh, "Frequency analysis of variable networks," Proc. IRE **38**, 291–299 (1950).
[40]C. Loeffler and C. Burrus, "Optimal design of periodically time-varying and multirate digital filters," IEEE Trans. Acoust., Speech, Signal Process. **32**, 991–997 (1984).
[41]P. Hoang and P. Vaidyanathan, "Non-uniform multirate filter banks: Theory and design," in Proceedings of the IEEE International Symposium on Circuits and Systems (1989), pp. 371–374.
[42]I. Djokovic and P. Vaidyanathan, "Results on biorthogonal filter banks," Appl. Comput. Harmon. Anal. **1**, 329–343 (1994).
[43]B. Edler and G. Schuller, "Audio coding using a psychoacoustic pre- and post-filter," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (2000), Vol. **2**, pp. 881–884.
[44]ISO/MPEG, "MPEG-4 Audio Version 2 ISO/IEC 14496-3:1999/Amd.1" (1999).
[45]R. Huber and B. Kollmeier, "PEMO-Q: A new method for objective audio quality assessment using a model of auditory perception," IEEE Trans. Audio, Speech, Lang. Process. **14**, 1902–1911 (2006).
[46]ITU-R Recommendation BS.1387-1, "Methods for objective measurements of perceived audio quality," International Telecommunication Union, Geneva (2001).
[47]B. C. J. Moore, *Cochlear Hearing Loss* (Wiley-Interscience, Malden, MA, 1998).
[48]L. Van Immerseel and S. Peeters, "Digital implementation of linear gammatone filters: Comparison of design methods," ARLO **4**, 59–64 (2003).
[49]S. Rosen and R. J. Baker, "Characterising auditory filter nonlinearity," Hear. Res. **73**, 231–243 (1994).
[50]T. Irino and R. Patterson, "A dynamic, compressive gammachirp auditory filterbank," IEEE Trans. Audio, Speech, Lang. Process. **14**, 2222–2232 (2006).
[51]E. Lopez-Poveda and R. Meddis, "A human nonlinear cochlear filterbank," J. Acoust. Soc. Am. **110**, 3107–3118 (2001).
[52]S. Mallat and Z. Zhang, "The matching pursuit software package (mpp)," ftp://cs.nyu.edu/pub/wave/software/mpp.tar.Z (Last viewed 4/23/2009).
[53]S. E. Ferrando, L. A. Kolasa, and N. Kovačević, "Algorithm 820: A flexible implementation of matching pursuit for gabor functions on the interval," ACM Trans. Math. Softw. **28**, 337–353 (2002).
[54]R. Gribonval and S. Krstulovic, "MPTK, The matching pursuit toolkit," http://mptk.gforge.inria.fr/ (Last viewed 4/23/2009).
[55]T. Herzke and V. Hohmann, "Improved numerical methods for gammatone filterbank analysis and synthesis," Acta. Acust. Acust. **93**, 498–500 (2007).
[56]M. Brucke, W. Nebel, A. Schwarz, B. Mertsching, M. Hansen, and B. Kollmeier, "Silicon cochlea: A digital VLSI implementation of a quantitative model of the auditory system," J. Acoust. Soc. Am. **105**, 1192 (1999).
[57]N. Ma, P. Green, and A. Coy, "Exploiting dendritic autocorrelogram structure to identify spectro-temporal regions dominated by a single sound source," Speech Commun. **49**, 874–891 (2007).

A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics

Muhammad S. A. Zilany
Department of Biomedical Engineering and Department of Neurobiology and Anatomy, University of Rochester, New York 14642 and Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario L8S 4K1, Canada

Ian C. Bruce
Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario L8S 4K1, Canada

Paul C. Nelson
Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland 21218

Laurel H. Carney
Department of Biomedical Engineering and Department of Neurobiology and Anatomy, University of Rochester, New York 14642

(Received 1 February 2009; revised 28 August 2009; accepted 1 September 2009)

There is growing evidence that the dynamics of biological systems that appear to be exponential over short time courses are in some cases better described over the long-term by power-law dynamics. A model of rate adaptation at the synapse between inner hair cells and auditory-nerve (AN) fibers that includes both exponential and power-law dynamics is presented here. Exponentially adapting components with rapid and short-term time constants, which are mainly responsible for shaping onset responses, are followed by two parallel paths with power-law adaptation that provide slowly and rapidly adapting responses. The slowly adapting power-law component significantly improves predictions of the recovery of the AN response after stimulus offset. The faster power-law adaptation is necessary to account for the "additivity" of rate in response to stimuli with amplitude increments. The proposed model is capable of accurately predicting several sets of AN data, including amplitude-modulation transfer functions, long-term adaptation, forward masking, and adaptation to increments and decrements in the amplitude of an ongoing stimulus.
 [DOI: 10.1121/1.3238250]

PACS number(s): 43.64.Bt, 43.64.Pg, 43.64.Wn [WPS] Pages: 2390–2412

I. INTRODUCTION

At the first synapse of the auditory pathway, the receptor potential of an inner hair cell (IHC) is converted into a discharge pattern on auditory-nerve (AN) fibers, where adaptation in discharge rate in response to a constant sound stimulus is observed. The IHC-AN synapse complex is believed to be mainly responsible for this adaptation. Although the mechanism that gives rise to synaptic adaptation is not completely understood, it could be caused either by the depletion of neurotransmitter from a readily releasable presynaptic pool of neurotransmitter (Moser and Beutner, 2000; Schnee *et al.*, 2005; Goutman and Glowatzki, 2007) or by the desensitization of post-synaptic receptors (Raman *et al.*, 1994).

Modeling the adaptation in the IHC-AN synapse has been a focus of extensive research over the last several decades. Early attempts employed a single-reservoir system with loss and replenishment of transmitter quanta (Schroeder and Hall, 1974; Oono and Sujaku, 1974, 1975), and later models added extra reservoirs (or sites) or more complex principles of transmitter flow control (Furukawa and Matsuura, 1978; Furukawa *et al.*, 1982; Ross, 1982, 1996; Schwid and Geisler, 1982; Smith and Brachman, 1982; Cooke, 1986; Meddis, 1986, 1988; Westerman and Smith, 1988). In general, the transmitter in these models lies in reservoirs or sites close to the presynaptic membrane and diffuses between reservoirs within the cell and out of the cell to the synaptic cleft. Each diffusion step is controlled by a permeability parameter, and at least one of the permeabilities is dependent on the stimulus. Mathematically, low-pass filters with appropriate orders and cut-off frequencies can replicate the replenishment and diffusion mechanisms between different transmitter reservoirs. Depending on the interconnection of the reservoirs, the flow of transmitter for these models can be implemented using either a cascade of low-pass filters or parallel low-pass filters.

Adaptation in the IHC-AN synapse is very complex. Its characteristics depend on stimulus intensity, duration, previous stimulation history, and spontaneous rate (SR) (Rhode and Smith, 1985; Relkin and Doucet, 1991). The diversity and complexity of adaptation pose a great challenge for successful modeling of the dynamics of this synapse. Two models with different structures have been developed independently in a series of studies (Meddis, 1986, 1988; Westerman and Smith, 1988; Carney, 1993; Zhang *et al.*, 2001; Sumner

et al., 2002, 2003). However, the mathematical descriptions of these two models are essentially equivalent despite their structural differences (Zhang and Carney, 2005). Both of these models are successful to some extent in simulating the onset adaptation responses (characterized by two exponential time constants) of the AN fibers. They have the same double-exponential adaptation (rapid and short-term) in both onset and offset responses (Zhang and Carney, 2005). However, physiological data exhibit substantially different dynamics between the offset and onset responses; in particular, the discharge rate may drop below the spontaneous rate at stimulus offset, sometimes to the point where there is a cessation of firing (i.e., the discharge rate is zero), followed by a relatively slow recovery to the spontaneous rate. Also the magnitude and time course of the onset and offset adaptations of the physiological data scale with the duration of the stimulus (Kiang, 1965), providing one illustration of the long-term behavior of AN response dynamics. Synapse models based on exponential adaptation fail to account for the offset adaptation as well as these long-term response properties. For example, physiological forward-masking data cannot be explained by these models without changing the model parameters such that they are inconsistent for onset and offset adaptations, and the dynamics must also be adjusted for fibers with different spontaneous rates (Meddis and O'Mard, 2005). These models also produce inaccurate responses to amplitude-modulated (AM) signals (Nelson and Carney, 2004; Zhang and Carney, 2005) and to increments and decrements in the amplitude of ongoing stimuli (Hewitt and Meddis, 1991).

Zhang and Carney (2005) developed a strategy that effectively avoids the constraint on the time course of recovery in the offset imposed by the onset parameters. A simple shift in the upward and downward directions (by same amount) of the pre-and post-synaptic responses, respectively, results in a slower recovery with a cessation in the post-synaptic response immediately after stimulus offset. It was reported that an appropriate shift can produce a better modulation transfer function (i.e., strength of AN synchronization to the envelope of amplitude-modulated stimuli as a function of modulation frequency) (Fig. 11, Zhang and Carney, 2005). However, such a shift also results in a systematic variation in the average rate with modulation frequency, which is not observed in AN responses (Joris and Yin, 1992), and also produces unrealistic steady-state rates of low spontaneous-rate fibers to tones at high sound levels (Nelson and Carney, 2004).

Hewitt and Meddis (1991) compared the responses of eight different synapse models to a set of standard stimuli and found no single model that could satisfactorily explain all of the data in their target set of responses. Although addition of extra reservoirs or sites in the model (equivalent to adding more exponential processes) tends to address more response properties of the AN (e.g., Smith and Brachman, 1982; Payton, 1988), such a model becomes mathematically intractable, and thus finding a set of parameters that works well for a large set of AN response properties is difficult, if not impossible.

Recently, power-law adaptation (PLA) has drawn a lot of attention in describing the dynamics of biological systems at levels ranging from single ion channels up to human psychophysics (Wixted and Ebbesen, 1997; Toib *et al.* 1998; Fairhall *et al.*, 2001; Leopold *et al.*, 2003; Ulanovsky *et al.*, 2004; Lundstrom *et al.*, 2008). Power-law adaptation is characterized by an adaptation of discharge rate that follows a fractional power of time or frequency rather than an exponential decay (Chapman and Smith, 1963). In fact, power-law dynamics can be approximated by a combination of a large number of exponential processes with a range of time constants (Brown and Stein, 1966; Thorson and Biederman-Thorson, 1974; Drew and Abbott, 2006). It has been argued that on short timescales, underlying mechanisms represent the contribution of intrinsic nonlinearities (e.g., ion channel dynamics). However, adaptation often exhibits power-law-like dynamics over longer timescales, implying the coexistence of multiple timescales in a single adaptive process (Camera *et al.*, 2006). In reality, multiple timescales exist in the multiplicity of channel dynamics present in a single neuron. To our knowledge, power-law dynamics has not yet been employed to explain adaptation at the level of the AN.

To illustrate a general model of power-law adaptation, suppose a stimulus $s(t)$ produces a response $r(t)$ that feeds back into an integrator $I(t)$, such that the adapted response, $r(t)=\max[0, s(t)-I(t)]$, and

$$I(t) = \alpha \int_0^t \frac{r(t')}{t - t' + \beta} dt' = \alpha r(t) * f(t)$$

$$\text{where } f(t) = 1/(t+\beta),$$

where α is a dimensionless constant and β is a parameter with units of time (Drew and Abbott, 2006). The suppressive effects of the response, $I(t)$, are accumulated with power-law memory that is intermediate between perfect (never forgotten) and exponential processes (Drew and Abbott, 2006). $I(t)$ can be described as a convolution of a power-law kernel, $f(t)$, with its prior responses, $r(t)$.

To compare the dynamics of adaptation between a power-law and a single exponential process, Fig. 1 illustrates power-law (solid) and exponential (dashed) adaptation in response to a unit step function $[s(t)=1,\ t>0]$ over four different time scales. For exponential adaptation [i.e., $I(t) = 1/\tau_a \int_0^t r(t')\exp((t'-t)/\tau_{ex})dt'$], the transient response decays exponentially to a steady-state value with a fixed time constant regardless of the stimulus time scale. Because the transition between the initial transient and the later sustained response occurs at a fixed time, exponential adaptation appears to have increasingly sharp transitions when observed over longer time scales. However, power-law adaptation has a similar shape for all four time scales, indicating the "scale-invariance" property of power-law adaptation. The responses appear qualitatively similar to exponential adaptation over any particular time period, and thus have no well-defined transient or sustained responses. Nevertheless, if a conventional time constant is evaluated from the responses of the power-law adaptation, its value depends on the duration of the responses being fit. This is illustrated by the power-law adaptation examples in Fig. 1, where the responses to the

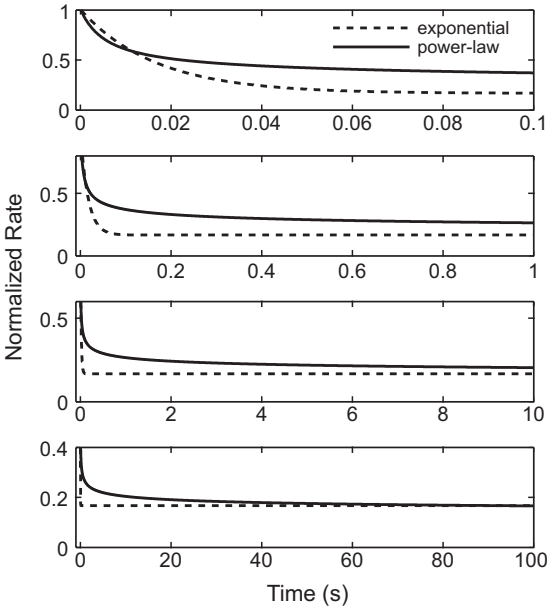


FIG. 1. Illustration of the dynamics of adaptation for exponential (dotted lines) and power-law (solid lines) models for four different time scales (0–0.1, 0–1, 0–10, and 0–100 s) in response to a unit step function. The parameters for the exponential adaptation are τ_a=0.2 s and τ_{ex}=0.1 s. The parameters for the power-law adaptation are $\alpha=5\times10^{-5}$ and $\beta=5\times10^{-3}$ s. The solid curves retaining a similar shape across different time scales demonstrate the "scale-invariance" property of the power-law adaptation.

four different stimulus durations appear to have transient responses of different durations, even though they arise from the same power-law adaptation.

Mathematically, the dimensionless constant α controls the amount of adaptation and hence makes the power-law adaptation scale-invariant (Drew and Abbott, 2006). In contrast, in the case of exponential adaptation, the equivalent of α has units of frequency ($1/\tau_a$, where τ_a is the time constant in seconds); thus, the transition between transient and sustained responses is fixed in time (i.e., it is not scale-invariant) (Drew and Abbott, 2006). Moreover, the long tail of the power-law kernel provides a longer memory for past responses than does exponential adaptation. The hypothesis of this study was that inclusion of power-law adaptation in the IHC-AN synapse could account for offset responses as well as other long-term response properties of the AN.

This paper describes a model of rate adaptation at the IHC-AN synapse that was incorporated into a composite phenomenological model of AN responses (Zilany and Bruce, 2006, 2007). Model responses were compared to physiological data for several different stimulus paradigms. The proposed PLA synapse model that includes both exponential and power-law dynamics replaces the previous synapse model having only exponential adaptation. Westerman and Smith's (1988) three-store diffusion model, which gives rise to exponential adaptation, is followed by two parallel power-law adapting paths that provide slowly and rapidly adapting responses, respectively. The parameters of the three-store diffusion model were adjusted to achieve desired onset responses with two time constants (rapid and short-term) and rate saturation at higher stimulus levels. It is worth mentioning that power-law adaptation alone does not result in rate saturation. The slowly adapting power-law component significantly improves the AN response at stimulus offset and also recovery after stimulus offset. The path with fast power-law dynamics contributes to the unsaturated onset response and to the "additivity" observed in AN rate responses to stimuli with amplitude increments. Several studies have confirmed that the process of short-term adaptation is additive in nature (Smith and Zwislocki, 1975; Smith, 1977; Abbas, 1979), meaning that the change in firing rate in response to an increment/decrement in stimulus level does not greatly depend on the time between the onset and the subsequent change in level. Smith *et al.* (1985) showed that this property also holds if increment responses are analyzed with different window lengths that separate the portions of the response associated with rapid and short-term adaptation. In contrast, the small-window decrement response decreases with increasing time delay (i.e., decrement responses are not additive over a short time window following the decrement). With the inclusion of power-law dynamics in the synapse model, the AN model presented in this paper can successfully account for a wide range of response properties of the AN, including additivity.

TABLE I. Parameter values.

Dynamics	Power-law adaptation α (dimensionless)	β (s)
Slow	5×10^{-6}	5×10^{-4}
Fast	1×10^{-2}	1×10^{-1}

Spontaneous rate (spikes/s)	Fractional Gaussian noise Standard deviation (spikes/s)
High (100)	200
Medium (5)	50
Low (0.1)	10

Another "long-term" property of AN responses that was addressed in this study is the pattern of correlations in response rates over long time intervals. The discharge rate of a single AN fiber is positively correlated over long time scales, whereas its response is often negatively correlated over the short term (Teich, 1989; Kelly *et al.*, 1996). Strong correlation of rate computed over widely separated analysis windows is referred to as "long-range-dependence" (LRD). Jackson and Carney (2005) investigated the implication of this effect of LRD in understanding SRs of AN fibers. They employed a fractional-Gaussian-noise-driven Poisson process to model LRD rates of AN fibers (Teich, 1989; Teich and Lowen, 1994). As LRD dramatically increases the variability of estimates of mean discharge rates (Jackson, 2003), they argued that the entire AN fiber population may be made up of neurons with only two or three true SRs. Incorporating appropriate LRD effects in their simulations, they successfully replicated the SR histograms of AN fibers. In order to model the distribution of SRs, the same approach was adopted in this study by adding a fractional Gaussian noise with appropriate parameters (Table I) in the slow power-law adaptation path of the IHC-AN synapse model.

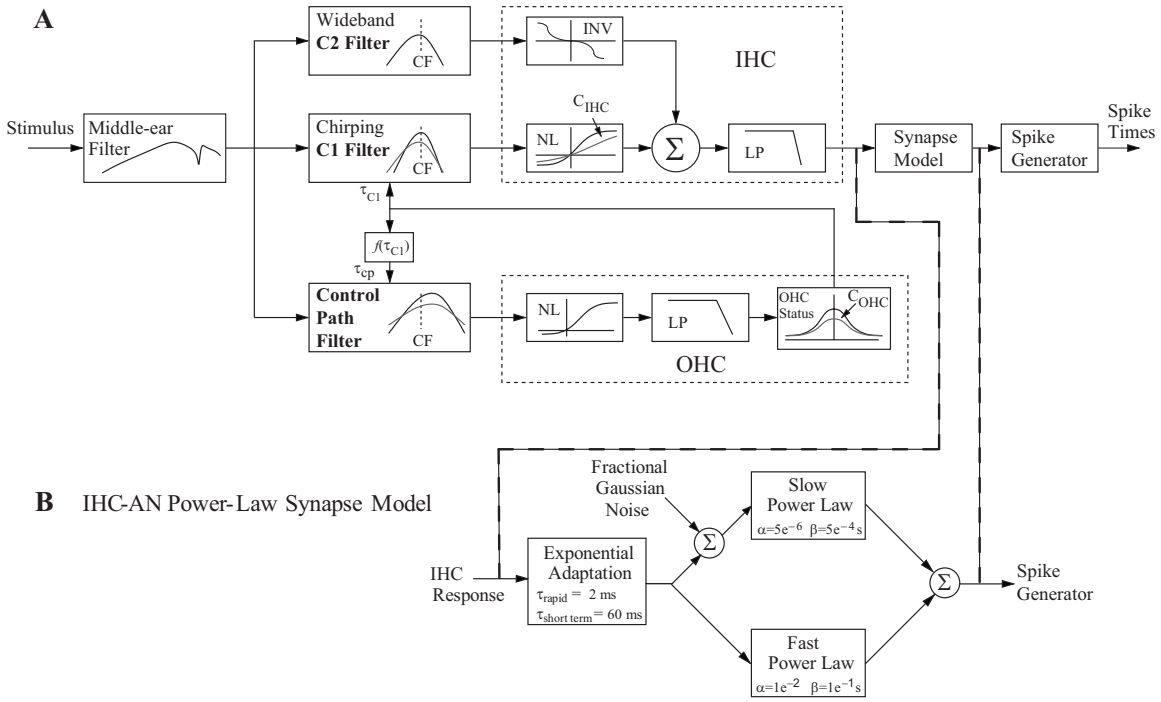


FIG. 2. (A) Schematic diagram of the model for the auditory periphery. The input to the model is an instantaneous pressure waveform of the stimulus (in pascals) and the output is a series of AN spike times. The model includes a middle-ear filter, a feed-forward control-path, a signal-path (C1) filter and a parallel-path (C2) filter, the IHC section followed by the synapse model, and the discharge generator. Abbreviations: outer hair cell (OHC), low-pass (LP) filter, static nonlinearity (NL), characteristic frequency (CF), and inverting nonlinearity (INV). C_{OHC} and C_{IHC} are scaling constants that specify OHC and IHC status, respectively. From Zilany and Bruce (2006, with permission). (B) IHC-AN synapse model: exponential adaptation (three-store diffusion model by Westerman and Smith 1987, 1988) followed by parallel power-law adaptation models (slow and fast). Fractional Gaussian noise added at the input of the slow power-law adaptation model results in the desired distribution of spontaneous rates.

II. DESCRIPTION OF THE MODEL

A. Architecture of the PLA model

A schematic diagram of the PLA model for auditory-nerve responses is shown in Fig. 2. Each section of the model provides a phenomenological description of the major functional components of the auditory periphery, from the middle ear (ME) to the auditory nerve. The input to the ME is an instantaneous pressure waveform of the stimulus (in pascals) sampled at 100 kHz. The ME filter is followed by three parallel filter paths: the C1 and C2 filters in the signal path and the broad-band filter in the control-path. The feed-forward control-path regulates the gain and bandwidth of the C1 filter to account for several level-dependent properties in the cochlea (Zhang *et al.*, 2001; Bruce *et al.*, 2003). The parallel-path C2 filter is implemented based on Kiang's two-factor cancellation hypothesis (Kiang, 1990). The combined response of the two transduction functions following the C1 and C2 filters provides the input to a seventh-order IHC low-pass filter (Zilany and Bruce, 2006, 2007). The IHC output drives the model for the IHC-AN synapse. In this study, a new model of the IHC-AN synapse replaced the previous synapse model. Finally the discharge times are produced by a renewal process that includes refractory effects (Carney, 1993). Detailed descriptions of the model stages are provided in Zilany and Bruce (2006, 2007); the model code is available at the following website: www.bme.rochester.edu/carney.

B. Modifications of the model from previous version

The model described and evaluated in this paper mainly differs from its predecessors (Zilany and Bruce, 2006, 2007) in the IHC-AN synapse section, which will be described in detail in the following sections. Another modification from the previous version of the AN model is that the cut-off frequency of the IHC low-pass filter was reduced from 3.8 to 3.0 kHz. The introduction of power-law adaptation in the synapse model significantly increases synchrony to pure tones, and thus the cut-off frequency was adjusted to match the maximum synchronized responses of AN fibers to pure tones as a function of characteristic frequency (CF) (Johnson, 1980).

It should be noted that in previous versions of the model, *responses of the synapse were simulated for only one repetition of the stimulus*. Because the discharge generator has relatively long-term dynamics that can span from one stimulus repetition to the next, a series of identical synapse output waveforms was concatenated according to the number of stimulus repetitions and the silent intervals between stimuli. In contrast, the PLA synaptic model presented here has power-law adaptation with memory that, in general, exceeds the duration of a single stimulus repetition. Thus, for the results described here, *the responses of the IHC model (not synapse) output were simulated for one repetition of the stimulus*, and then a series of identical IHC responses was concatenated and used as the input to the synapse model.

C. PLA model of the IHC-AN synapse

Although many biological systems exhibit power-law rather than exponential dependence on time, in some cases, power-law adaptation alone underestimates the amount of adaptation at short-times (Drew and Abbott, 2006). For example, the response of an electrosensory neuron in electric fish to a long duration (100-s) amplitude-modulated step stimulus (Xu *et al.*, 1996) was well described by power-law adaptation from 20 ms to 100 s, but not from 0 to 20 ms [Fig. 2(b), Drew and Abbott, 2006]. This observation led Drew and Abbott (2006) to argue for the presence of an additional exponential adaptation component with a small time constant. It is well-known that adaptation to sustained tones in mammalian AN fibers involves at least three time scales: rapid adaptation on the scale of milliseconds, short-term adaptation on the scale of several tens of milliseconds, and slow adaptation on the scale of seconds (Kiang, 1965). In order to include all of these time scales, the new IHC-AN synapse model has power-law adaptation following short-term exponential adaptation components.

The variation in adaptation characteristics across different AN fibers suggests that individualized sets of model parameters might be required to predict individual AN fiber responses accurately. However, the goal of this study was to determine a single parameter set that was satisfactory for a wide range of response properties of AN fibers.

1. Exponential adaptation

This part of the synapse model is exactly the same as in previous versions of the model (Zhang *et al.*, 2001; Zilany and Bruce, 2006, 2007), which included a time-varying implementation of Westerman and Smith's (1988) three-store diffusion model. The parameters were determined according to the derived equations (Appendix A of Westerman and Smith, 1988) based on the desired response characteristics for the onset and steady-state responses of the post-stimulus time histograms (PSTHs) to tones (Appendix in Zhang *et al.*, 2001).

The onset response of the model AN fiber is governed by exponential adaptation with two time constants (2 and 60 ms). The other parameters of the three-store diffusion model in the exponential adaptation stage were set to produce spontaneous activity and rate saturation at higher stimulus levels (Zhang *et al.*, 2001).

2. Power-law adaptation

In the PLA model, the output of the exponential process drives two parallel power-law adaptation paths, namely, slow and fast power-law adapting components. The inclusion of two power-law functions in the model was motivated by the fact that one power-law adaptation component alone cannot account for an important AN response property, additivity (see below), while retaining the onset adaptation dynamics set by the exponential processes. The selection of parameters for these two power-law functions is more challenging and was complicated by the fact that power-law adaptation has no well-defined transient or sustained responses (Fig. 1). So, rather than trying to fit individual data sets, parameters of the power-law functions were chosen in such a way that the model qualitatively addressed a range of AN response properties for a wide variety of stimulus conditions. The parameters were then kept fixed and were not optimized to fit individual AN responses.

The parameters of the slow power-law component were such that it closely followed the output of the exponential adaptation model for onset responses (i.e., slow power-law adaptation further adapts the signal, but with a time course that is similar to that of its input). Because the power-law has longer memory than the exponential function, the offset and other long-term response properties were significantly improved in the output of the slow power-law component. Thus, model predictions for forward-masking paradigms and for amplitude-modulated signals were also improved substantially by inclusion of the slow power-law adaptation component.

However, the desired property of AN additivity cannot be modeled with a power-law function that has the same time course of adaptation as the exponential adaptation (see below). To capture the phenomenon of additivity, a second power-law function with faster adaptation was therefore introduced in the model; this function adapts quickly and is very responsive to increments in amplitude of an ongoing stimulus. Thus the change in discharge rate in response to an increment remains almost the same irrespective of the delay between stimulus onset and presentation of the increment. However, in response to decrements, both power-law components turn off instantaneously and recover very slowly. As the fast power-law component is very sensitive to increments of the stimulus, it results in a highly synchronized response to the envelope of amplitude-modulated signals and also to pure tones at low frequencies (this synchrony is limited by the IHC low-pass filter).

As stated earlier, the parameters of the power-law functions were adjusted to qualitatively address a wide range of response properties of the AN. To determine the parameters of the slow power-law function, two particular data sets were used that require adaptation with longer memory and thus were relevant to the power-law dynamics. The first one was the offset responses to a pure tone stimulus across several sound levels (Kiang, 1965), and the other one was the responses to a probe in a forward-masking stimulus paradigm (Harris and Dallos, 1979). Once the parameters for the slow power-law component were set, the parameters of the fast power-law function were then chosen by qualitatively matching the model responses with the physiological data for the increment/decrement paradigm (Smith *et al.*, 1985). The parameters of both slow and fast power-law functions are provided in Table I. After all parameters were set, the model was tested for a wide variety of AN response properties; the results are reported in Sec. III.

3. Implementation of the power-law function

The computation of power-law functions is very expensive.[1] As the duration of the signal increases, the corresponding computational time increases significantly because computation of each sample of the adapted response requires memory from the onset of the signal (i.e., onset of

the first repetition in case of more than one repetition of the signal, which corresponds to time zero). As mentioned earlier, the power-law function can be expressed as the convolution of power-law kernel with its prior responses. When possible, for computational efficiency, power-law kernels of fast and slow power-law functions were approximated by sixth- and tenth-order infinite impulse response (IIR) filters, respectively. To ensure stability, these digital filters were implemented as a cascade of second-order systems. The responses of the model for actual and approximate implementations were almost the same for short duration stimuli (Fig. 5). However, for very long stimuli (as in Fig. 6), the actual implementation of the power-law functions was required to replicate the physiological data.[2]

4. Implementation of SR

To model the distribution of SR, the fractional Gaussian noise (fGn) was added in the slow power-law adaptation path of the synapse model. The source of this noise within the auditory periphery is not known; it was introduced in the slow power-law path of the model for the following reasons. First, the parameters of the slow power-law path did not alter the dynamics of the noise significantly, whereas both exponential and fast power-law adaptation would have changed the noise dynamics substantially. That is, fGn maintains the spectral properties of $1/f$ type noise with slightly altered magnitude after the slow power-law adaptation. Note that the fluctuation in the fGn also prevents the spontaneous rate from continuously adapting toward a value of zero (result not shown). Second, if the noise were added directly to the synapse output, the added noise would "fill in" the pause in the offset responses, and thus the dynamics of recovery would be obscured by the noise.

Three parameter sets were used in this study to generate fGn (with Hurst index $H=0.9$, which specifies the strength of the LRD) corresponding to three classes of SR (low, medium, and high). The rationale behind employing three true SRs rather than two (a possibility suggested by Jackson and Carney, 2005) will be examined in detail in Sec. IV. These parameters, provided in Table I, were adjusted to simulate the distribution of AN SRs in cat (Liberman, 1978). Because the exponential adaptation model has a steady-state response that determines spontaneous rate, the added fGn has zero mean. It is worth noting that these parameters are different from those used in Jackson and Carney (2005) for two reasons. First, in Jackson and Carney (2005), refractory effects were not included in the Poisson process, whereas the discharge generator in the PLA model has refractory effects to simulate realistic responses of the AN. Second, the parameter values compensate for the slight alteration of the noise dynamics by the slow power-law adaptation.

III. RESULTS

In this section, the spontaneous activity as well as responses of the model to a wide variety of stimuli, including paradigms involving tones and noise, are compared to physiological data from the literature.

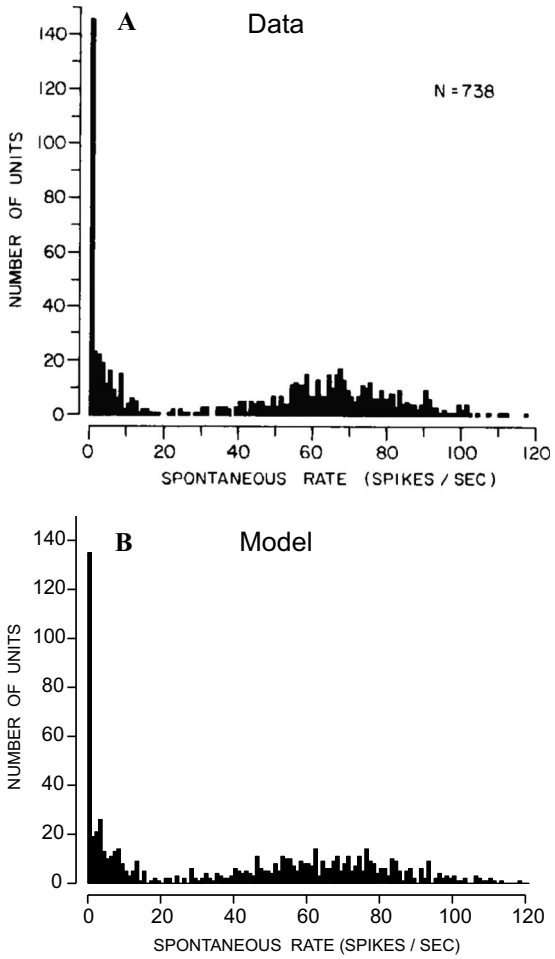


FIG. 3. Histogram of actual (upper panel) and model (lower panel) SR estimates from 30-s recordings from 738 fibers in the auditory nerves of cats (binwidth of 1 spike/s). (A) Actual AN SR histograms from Liberman (1978, with permission). (B) Model histogram of SR estimates using the same paradigm as in Liberman (1978) for 738 independent simulations. Three parameter sets for the fGn were used, applied to the proportions of different SR fibers reported in Liberman (1978). fGn parameters are provided in Table I.

A. Spontaneous activity

The upper panel (A) in Fig. 3 is a histogram of SR estimates from 30-s recordings for 738 cat AN fibers (Liberman, 1978). The lower panel (B) is a histogram of model SR estimates, made using a paradigm that matched Liberman's (1978). A total of 738 independent simulations was carried out, with the number of simulations for each SR class determined according to the proportions of different SR fibers reported in Liberman (1978) [high SR (~61%), medium SR (~23%), and low SR (~16%)]. As described in Sec. II, each SR type was simulated by choosing one of three possible parameter values for the fGn (Table I). The model histogram matches the distribution of SRs reported for the physiological data.

B. Responses to pure tones at CF

1. Recovery of spontaneous activity

At the offset of a tone pip, AN firing can be substantially reduced relative to spontaneous rate and is often characterized by a pause in the response followed by a slower recov-

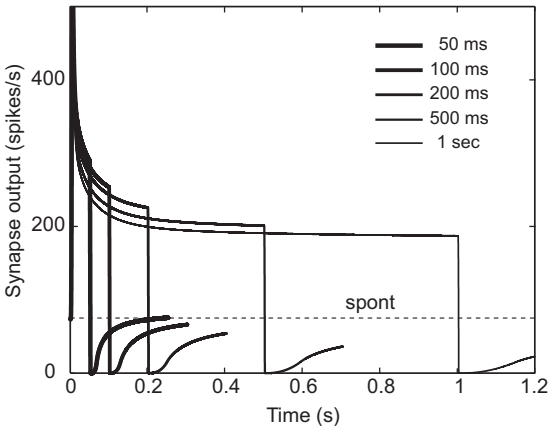


FIG. 4. Illustration of the "scale-invariance" property of the PLA model. Here the output of the synapse model is shown before the discharge generator. The fGn was not included in the model to avoid fluctuation in the output. The dotted line indicates the spontaneous rate of the fiber. The stimulus was a 10-kHz tone at CF, 12 dB above threshold. The duration of the signal varied from 100 ms to 1 s, but the inter-stimulus interval was fixed at 200 ms in all cases. Responses to 50 repetitions of the stimulus were averaged. The dynamics of recovery from the stimulus offset to spontaneous rate scales according to the duration of the signal.

ery (on the order of several tens of milliseconds) to spontaneous activity (Harris and Dallos, 1979; Smith, 1977; Westerman, 1985). The amount of reduction in rate and the exact nature of recovery depend on the stimulus level (Yates *et al.*, 1985) and also on the fiber's spontaneous rate (Relkin and Doucet, 1991). Low spontaneous-rate (LSR) neurons take a longer time to recover from prior stimulation as compared to high spontaneous-rate (HSR) neurons (Relkin and Doucet, 1991).

To demonstrate the scale-invariance property of the PLA model, the output of the synapse model is shown in Fig. 4 in response to a tone stimulus (tone at CF = 10 kHz, 12 dB above threshold) with different durations, but with a fixed inter-stimulus interval of 200 ms. The signal durations used were 50, 100, 200, 500, and 1000 ms. Responses to 50 repetitions of the stimulus were averaged. The dotted line indicates the (high) spontaneous rate of the fiber. To avoid fluctuations in the output and to emphasize the relevant response details for this simulation, fGn was not included in the model for this illustration. For short-duration signals (<~200 ms), a 200-ms silent interval is adequate for full recovery to spontaneous rate, whereas longer signals require longer inter-stimulus intervals to completely recover to spontaneous rate. Since power-law adaptation has long memory for past responses, the dynamics of recovery after signal offset for the PLA model scales according to the duration of the signal. In contrast, the recovery to spontaneous rate in the exponential adaptation model (results not shown) would occur over a constant time period irrespective of the duration of the signal because the time constant of the exponential process is fixed. It should be noted that the relatively steady-state part of the PLA model response is noticeably reduced in response to longer duration signals because the responses do not fully recover to spontaneous rate during the inter-stimulus interval.

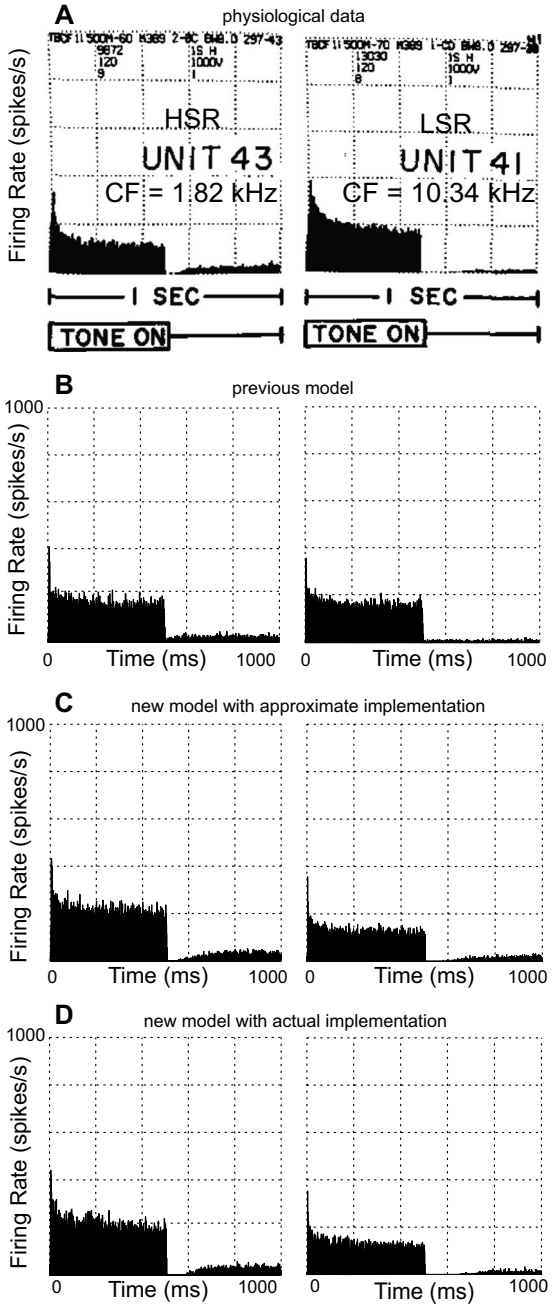


FIG. 5. Effects of spontaneous rate on recovery in experimental [upper panels (A)] and model [lower panels (B)–(D): previous model, new PLA model with approximate and actual implementations, respectively] histograms of two AN fibers in response to 500-ms duration constant-amplitude stimuli. The stimuli were presented once a second. Each histogram represents 2 min of data collection. Left panels: CF = 1.82 kHz, HSR (unit 43 in data); right panels: CF = 10.34 kHz, LSR (unit 41 in data). (A) From Kiang (1965, with permission). (B) Model histograms of AN fibers at 25 dB SPL using the previous model (Zilany and Bruce, 2007) that has only exponential adaptation in the synapse model. (C) PLA model histograms at 25 dB SPL with approximate power-law implementation. (D) PLA model histograms at 25 dB SPL with actual power-law implementation.

Figure 5(a) shows PSTHs for a HSR AN fiber with CF = 1.82 kHz on the left and for a LSR AN fiber with CF = 10.34 kHz on the right (from Kiang, 1965). The stimulus was 120 repetitions of a 500-ms tone followed by a 500-ms silent period. Figure 5(b) shows corresponding responses of the previous AN model that had only exponential adaptation in

the synapse model. In contrast to the physiological data, the responses of the previous model show no pause in the response after the stimulus offset and a very quick recovery to spontaneous activity. The two lower panels [Figs. 5(c) and 5(d)] show the PLA model responses for both the approximate and actual implementations of the power-law functions. In general, the PLA model responses closely resemble the physiological data. Also, as expected, the model response computed using the actual power-law implementation has a slightly slower recovery than the response computed with the approximation.

2. Long-term recovery

Young and Sachs (1973) measured the recovery of the discharge rate of single AN fibers to tone pips after exposure to 60-s long continuous tones. Both the exposure and test tones were at the fiber's CF. The 100-ms test tones were presented once per second at 19 dB sound pressure level (SPL) either before or after the exposure. The total duration of pre- and post-exposure test signals were 10 and 60 s, respectively. Effects of exposure level on recovery were studied at four exposure SPLs (29, 59, 74, and 89 dB). The post-exposure test-tone response rates were fitted to an exponential to determine the time constant of recovery.

Figure 6 shows the recovery of post-exposure responses (to pre-exposure response rates) for a HSR AN fiber with CF = 2.15 kHz, using the stimulus paradigm described above. The left panels [(A)and (C)] show physiological responses from cat (Young and Sachs, 1973), and the right panels [(B) and (D)] show corresponding model responses. Recovery of the post-exposure response was fitted to an exponential, and the computed time constants are shown in the lower panels. The stimulus paradigm was the same for both actual and model fibers, except that the test signal used for the model was reduced to 9 dB SPL, to approximately match the level with respect to threshold to that of the cat AN fiber. Model responses to ten repetitions of the input stimulus were averaged, as was done for the experimental data.

Following exposure, the discharge rate to the test tone is transiently reduced, and the time constant of recovery increases as the exposure level increases, even though responses during the exposure saturate in response to higher-level exposure tones. Young and Sachs (1973) argued that there exists an additional suppression mechanism other than exposure evoked suppression to account for this phenomenon. The PLA model with two parallel power-law adaptation paths can qualitatively address this issue. Although the steady-state rate saturates at higher levels, model responses at onset have a much wider dynamic range (Smith, 1988). As the power-law function has a long memory (which extends back to the onset of the exposure stimulus), the reduction in the test signal responses continues to increase for higher-level exposure tones. In addition to the slow power-law component, the fast power-law component also plays a significant role in this case, as this component is very sensitive to level at the onset of the stimulus. For a good quantitative fit between model responses and actual data, the two parallel power-law adaptation paths could be driven by two separate inputs with a significant emphasis on the fast power-law component (results not shown). As the mechanism of these power-law functions is not known, and to keep the model structure simpler, both power-law functions of the PLA

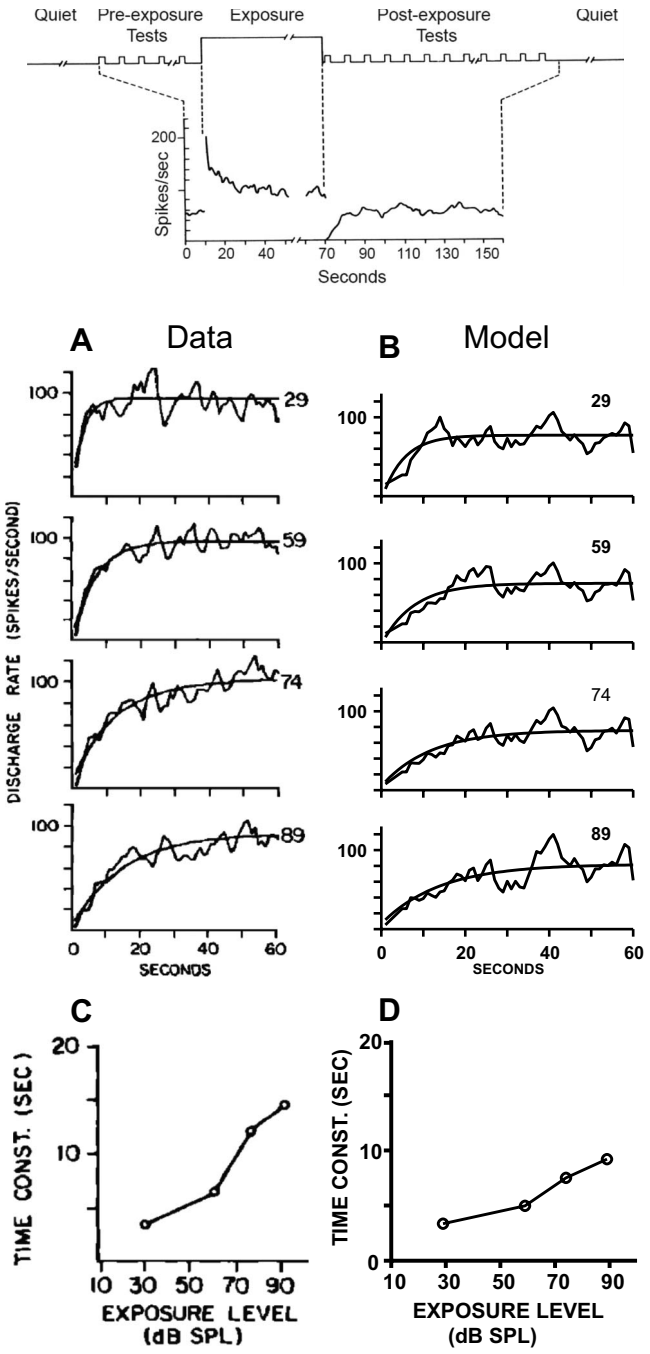


FIG. 6. Effect of exposure level on recovery for an AN fiber with CF = 2.15 kHz (HSR). The stimulus paradigm is illustrated at the top. Left panels [(A) and (C)] show the actual experimental responses, and the right panels [(B) and (D)] show the corresponding model responses. Duration of the exposure signal was 60 s, and the exposure levels were 29, 59, 74, and 89 dB SPLs (shown on the right/above of each curve). The test signal (100-ms long applied once per second) was also at CF with level 19 dB SPL (fixed). However, the test signal level in the model responses was at 9 dB SPL to match with the level of the experimental fiber with respect to its threshold. Total durations of the pre- and post-exposure test signals were 10 and 60 s, respectively. Recoveries of the post-exposure responses (fitted to an exponential) are shown with their corresponding time constant values. [(A) and (C)] From Young and Sachs (1973, with permission). [(B) and (D)] Model responses of recovery employing the same experimental condition as in the data. Responses to ten repetitions of the same stimulus were averaged.

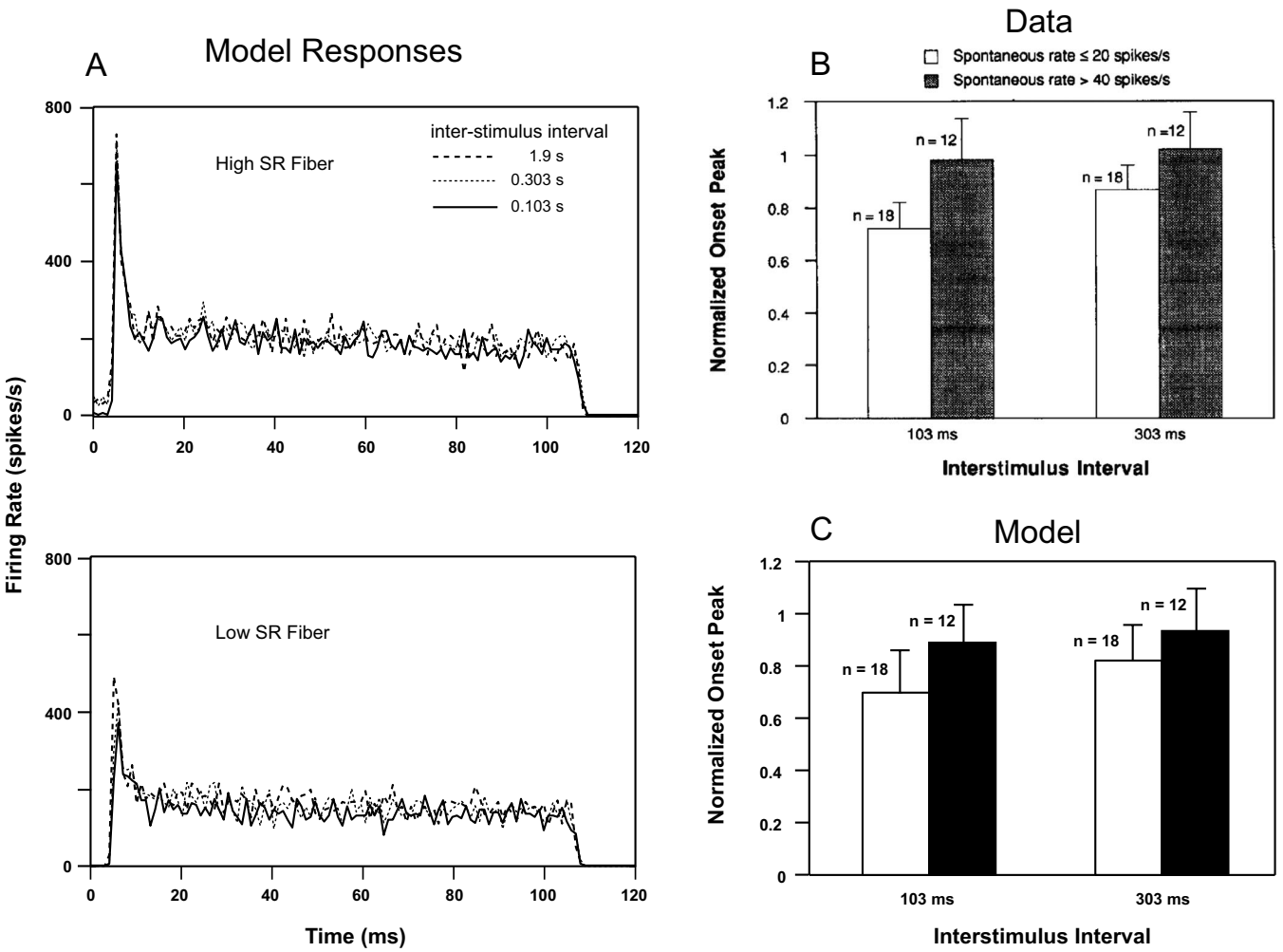


FIG. 7. Effect of inter-stimulus intervals on the responses of high and low SR fibers. (A) PST histograms of the PLA model in response to a 100-ms signal (tone at CF = 2 kHz), 40 dB above threshold for a HSR (upper panel) and a LSR (lower panel) fiber. The inter-stimulus intervals were 1.9, 0.303, and 0.103 s. With decreasing inter-stimulus interval, the peak of the onset was more reduced for the LSR fiber than for the corresponding HSR fiber. [(B) and (C)] Averaged value of normalized magnitude of the onset peak of PST histograms vs inter-stimulus intervals (0.103 and 0.303 s) for high (solid bar) and low (open bar) SR fibers is shown. The onset peak for each neuron was normalized by the onset peak of that neuron when the inter-stimulus interval was 1.9 s. The normalized values were averaged for all neurons within a group for different inter-stimulus intervals. (B) From Relkin and Doucet (1991, with permission). (C) Model responses from AN fibers with CFs ranging from 1 to 20 kHz for both high and low SRs.

model are driven by the same input (i.e., the exponentially adapted IHC output). Note that since the recovery in the responses of the previous models (that have only exponential adaptation) does not scale with the duration of the stimulus, those models cannot account for these long time constants (on the order of several seconds) of recovery.

3. Effects of SR and inter-stimulus interval on adaptation at tone onset

Rhode and Smith (1985) and Müller and Robertson (1991) investigated the effect of fiber types on adaptation after stimulus onset in cat and guinea-pig, respectively. They found that LSR fibers show no or very little adaptation, whereas HSR fibers show substantial adaptation. However, Relkin and Doucet (1991) pointed out that the inter-stimulus intervals used in these studies may have been too short to allow for full recovery from stimulation in previous repetitions, especially for LSR fibers. They reported that an inter-stimulus interval of 300 ms was long enough for a 100-ms duration signal (40 dB above threshold) to allow onset responses to fully recover in HSR fibers, but not in LSR fibers. The model was used to simulate their experiment using both HSR and LSR model fibers.

Left panels in Fig. 7(a) show the PST histograms of the PLA model for a HSR (upper) and a LSR (lower) fiber. The tone stimulus at CF (2 kHz) was 100 ms in duration with a level 40 dB above threshold and was presented 50 times with inter-stimulus intervals of 0.103, 0.303, and 1.9 s. For the LSR fiber, the peak at the stimulus onset was reduced to ~74% with decreasing inter-stimulus interval from 1.9 to 0.103 s. In contrast, the peak onset of the HSR fiber was decreased to only ~90% for the same condition. It should be noted that the spontaneous rate of the HSR fiber was significantly reduced for the 0.103-s inter-stimulus interval condition because the duration of the inter-stimulus interval was not sufficient to allow full recovery to spontaneous rate before the onset of each subsequent signal (see Fig. 4). For the results in the right panels, the onset spike rate was computed using the number of spikes in the most populated 1-ms bin of the response histogram after the onset of the stimulus. The

onset peak for each AN fiber was normalized by the onset peak for that neuron when the inter-stimulus interval was 1.9 s. The right upper panel [Fig. 7(b)] shows the averaged responses from 12 HSR and 18 LSR AN fibers from chinchilla (Relkin and Doucet, 1991). Solid bars show the responses of HSR fibers, and open bars represent LSR fiber responses. Model responses shown in Fig. 7(c) were also averaged from AN fibers with CFs ranging from 1 to 20 kHz (logarithmically spaced) for both HSR and LSR fibers. Similar to the physiological data, the model HSR fibers were almost completely recovered when the inter-stimulus interval was 0.303 s, but the LSR fibers were ~80% recovered by this time. This result is further supported by the observation of Young and Sachs (1973) that different SR classes have different time course of recovery at equal sound levels. However, their behavior was identical for SR classes when plotting the recovery of time constants vs driven rate instead of stimulus level. The PLA model responses are also consistent with the observation by Young and Sachs, 1973 (results not shown).

C. Responses to tones with amplitude increments/decrements

1. Conservation of energy

Westerman and Smith (1987) reported that the total transient response associated with an incremental stimulus paradigm shows a form of conservation. They computed the transient AN responses for two contiguous 300-ms tone bursts with the first tone (at CF) varying in level (5, 10, 15, and 20 dB above threshold) and the second tone (also at CF) fixed at a higher level (43 dB above threshold). Transient response components were obtained by fitting the histograms to a characteristic equation [having rapid, short-term, and sustained responses (Westerman and Smith, 1987)]. Then component integrals were calculated separately from the background (first tone) and increment portion of the response histogram. The integral of each component is the product of the component magnitude and the time constant and equals the number of spikes contributed by that component to the total transient response.

The upper panels [(A) and (B)] in Fig. 8 show the PST histograms of one AN fiber (CF = 5.99 kHz, HSR) in response to the above incremental stimulus paradigm. Panel (A) represents the physiological response from a gerbil AN fiber (Westerman and Smith, 1987), and panel (B) shows the corresponding model responses. As the level of the first (so-called "background") tone increases, the amount of transient response associated with it also increases, whereas the transient activity in response to the second tone decreases.

The rapid and short-term transient components were evaluated separately for both background and increment portions of the tone and are shown in the lower panels [(C) and (D)]. Panel (C) shows the average results for seven gerbil AN fibers (Westerman and Smith, 1987). Panel (D) represents the model's rapid and short-term components determined from the model histograms shown in panel (B). The combined transient response associated with the two portions of the stimulus (background and post-increment) remains roughly constant (although slightly less for the rapid component) and thus exhibits conservation.

2. Increments/decrements

The effects of prior adaptation on responses in the increment/decrement paradigm are illustrated in Fig. 9. Left panels (A) compare the increment responses of the model AN fiber with three different versions of the synapse model: one with only exponential adaptation (i.e., the previous model), one with exponential followed by slow power-law adaptation (middle panel), and one with exponential followed by both slow and fast power-law adaptations (lower panel). The stimulus was a 60-ms duration pedestal tone at CF (4.16 kHz), 13 dB above threshold, with a 6-dB increase in level occurring at various delays (up to 40 ms) after the onset of the pedestal. The increment responses were obtained by subtracting the response to the pedestal tone from the response to the tone with an increment in level. It is evident that the responses to the increment paradigm were not additive in the first two cases (especially for the short window at the onset of the increment), which justifies the inclusion of a fast power-law component in addition to a slow power-law adaptation component in the PLA model (as mentioned in Sec. II C 2). In Fig. 9(b), the change in firing rate was analyzed over two windows: 0.64 ms (onset window, circles) and 10.2 ms (large window, upward triangles), both windows beginning at the time of the change in the response following the increment. Dotted lines show the physiological data from gerbil (Smith *et al.*, 1985), and the solid lines represent the corresponding PLA model responses (of a HSR fiber) for the same stimulus paradigm. For both physiological data and model responses, the incremental change in discharge rate remains almost constant irrespective of the delay. As mentioned earlier, the fast power-law component in the PLA model adapts very quickly and is also very sensitive to increments of the stimulus. As a result, the change in discharge rate to an increment in stimulus level is almost the same, irrespective of the time delay at which the increment occurs. Thus, model responses exhibit additivity for both small and large analysis windows in response to increments in tone level.

Figure 9(c) shows the change in firing rates of an AN fiber (CF = 3.58 kHz, HSR) for decrements in level to an ongoing stimulus. The decrement stimulus paradigm is similar to the increment paradigm, except that the change in level is negative. Both small (0.64-ms) and large (10.2-ms) analysis windows were used (circles and upward triangles, respectively). Dotted lines show the physiological data from Smith *et al.* (1985), and solid lines indicate the corresponding PLA model responses. As in the physiological data, model responses after decrements are additive for the large window analysis, but onset window decrements are clearly not additive.

D. Forward masking

The responses of AN fibers to a probe stimulus are reduced immediately following stimulation by a masker. This

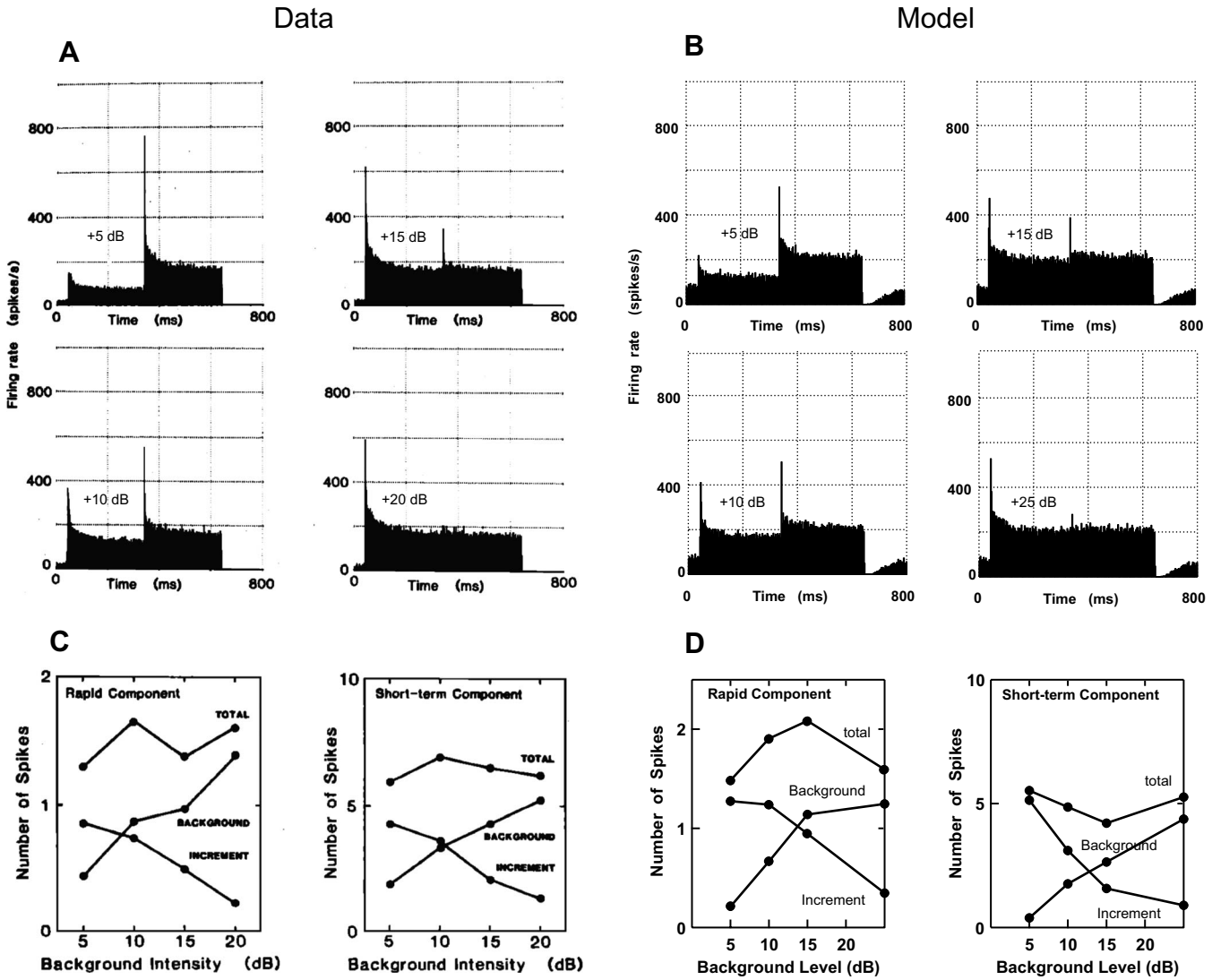


FIG. 8. AN fiber histograms and conservation of adaptation in both rapid and short-term components for the amplitude increment response paradigm (binwidth of 2 ms). [(A) and (C)] Physiological responses from gerbil (Westerman and Smith, 1987). [(B) and (D)] PLA model responses. The stimulus was at CF (5.99 kHz, HSR) with duration of 600 ms. The initial levels of the tone were 5, 10, 15 and 20 dB above threshold (background). At 300 ms, the intensity was increased to 43 dB above threshold (increment) in all cases. (A) Actual AN fiber histograms from Mongolian gerbil: from Westerman and Smith (1987, with permission). (B) Model histograms using the same paradigm as above, except that the highest level of the background tone was 25 dB above threshold (because the model fiber shows a wider dynamic range than the corresponding AN fiber of the physiological data). (C) Mean values of rapid and short-term components from six fibers; from Westerman and Smith (1987, with permission). (D) Model transient responses (for one AN fiber) from the corresponding model histograms (one fiber) shown in (B) using the same method as employed in the data.

reduction in response is presumed to be a function of adaptation and is likely to contribute to the psychophysical phenomenon of forward masking. Several physiological studies have been performed in different species to study the recovery of AN responses using forward-masking paradigms (e.g., Smith, 1977; Harris and Dallos, 1979; Westerman, 1985).

Figure 10 shows an example of the post-stimulus recovery function of a chinchilla AN fiber (CF = 2.75 kHz, HSR) in the left panels (Harris and Dallos, 1979), and the model responses with the same paradigm are shown in the right panels (B). The masking stimulus was 100 ms in duration, tone frequency was matched to CF (2.75 kHz), and tone level was 30 dB above threshold. The probe was 15 ms in duration, 20 dB above threshold, and its frequency was matched to CF. The probe responses are expressed as a percent of the control response (i.e., when there was no masker) and are shown as a function of probe delay, ranging from 1 to 150 ms. The histograms on the right show the responses that were used to compute the data points on the left. The PLA model responses agree with the physiological data; as the delay between masker offset and probe onset increases, the probe responses are less reduced as the AN fiber shows more recovery from adaptation. In contrast, the previous model shows significantly less reduction in rate than the physiological data, especially at short delays [shown by the dotted line, Fig. 10(b)]. In fact, the masked probe response of the previous model never fell below 50% of the control response, even at very small delays and high masker levels (result not shown).

The influence of masker level on the post-stimulus recovery function is shown in Fig. 11. The same paradigm described above was used, except that the masker level var-

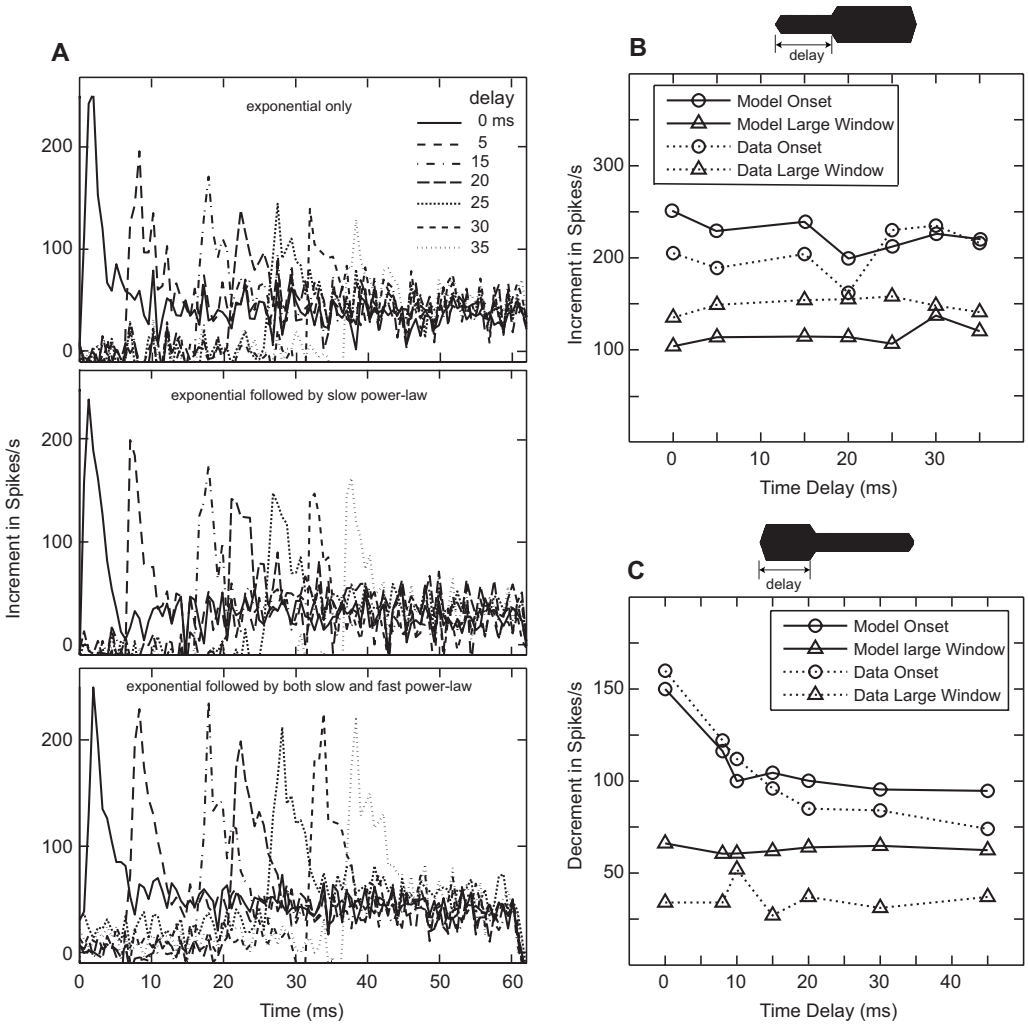


FIG. 9. Effects of prior adaptation on increment and decrement responses. (A) Model increment responses as a function of time for three different conditions of the synapse model (with only exponential adaptation, exponential followed by slow power-law adaptation, and exponential followed by both slow and fast power-law adaptations). (B) Increment responses from both actual and PLA model responses. (C) Decrement responses. The stimulus was a 60-ms CF tone 13 dB above threshold, and subsequently levels were either increased or decreased by 6 dB at different delays from onset. The resulting increment/decrement responses are obtained by subtracting the response to the constant-intensity tone (pedestal) from the response to the tone with an increment/decrement in level. Changes in rate responses for both the onset window (first 0.64 ms, circles) and a larger window (first 10.2 ms, downward triangles) are shown. Dotted lines with symbols show the actual data (data points from Figs. 5 and 7 of Smith *et al.*, 1985), and the solid lines with symbols represent the corresponding PLA model responses. (B) Increment responses. CF at 4.16 kHz (HSR). (C) Decrement responses. CF at 3.58 kHz (HSR).

ied from 10 to 60 dB above threshold. For both experimental and model paradigms, an inter-masker interval of 230 ms was used for the +10 and +20-dB maskers but was increased to 330 ms for higher masker levels to minimize the buildup of long-term effects. The upper panel (A) shows the median responses from 37 fibers with CFs ranging from 0.5 to 16 kHz from chinchilla (Harris and Dallos, 1979). Model responses shown in the lower panel (B) are averaged from ten fibers (six HSR and four LSR fibers) with CFs spaced logarithmically across the same range. Both the time course of recovery and the magnitude of forward masking increase with increasing masker level, and both tend to saturate at higher masker levels. Although PLA model responses qualitatively match with the chinchilla data, the recovery of model probe responses in the mid-delays (10–50 ms) is greater than the corresponding physiological responses; this difference could possibly be explained by differences between the spontaneous rates of the model and the data which are not specified in Harris and Dallos (1979). It should be noted that model LSR fibers show longer time courses of recovery and more reduction in probe response than the corresponding responses of HSR fibers.

E. Responses to amplitude-modulated tones

A systematic study of cat AN responses to sinusoidally amplitude-modulated (SAM) tones by Joris and Yin (1992) serves as an excellent template for a detailed evaluation of the PLA model in response to AM stimuli. The equation representing a SAM signal is given by

$$s(t) = [1 + m \sin(2\pi f_m t)]\sin(2\pi f_c t),$$

where m is the modulation depth and f_m and f_c are modulation and carrier frequencies, respectively. Figure 12 illustrates the effect of increasing modulation depth (m) on PSTH shapes and the corresponding synchrony and modulation gain of an AN fiber with CF = 20.2 kHz (HSR fiber). The left panels [(A) and (C)] show the physiological responses

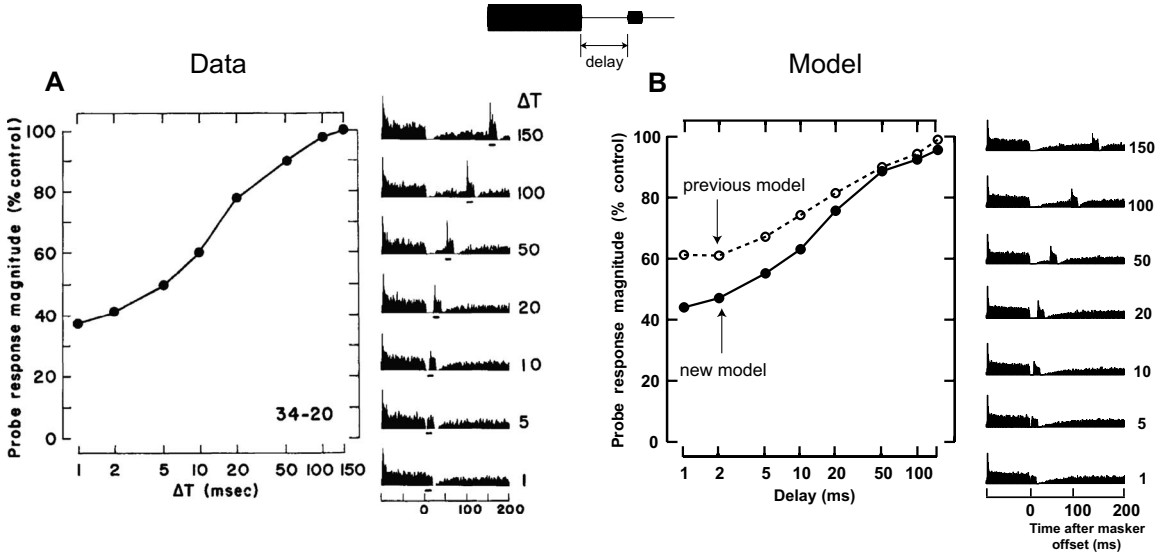


FIG. 10. Actual [left panels (A)] and model [right panels (B)] post-stimulus recovery as a function of delay between masker offset and probe onset. Masker: 2.75-kHz tone (fiber's CF, HSR), 30 dB above threshold (+30 dB), 100-ms duration. Probe stimulus: 2.75-kHz tone, +20 dB, 15-ms duration. Each data point represents the average number of spikes evoked by the probe as a percent of the control response (probe alone). The PST histograms on the right are the source for the data points on the left. (A) From Harris and Dallos (1979, with permission). (B) Model responses for the same paradigm as in the experiment. The solid line with filled circles represents the responses of the PLA model, and the dashed line with open circles indicates the responses of the previous model (Zilany and Bruce, 2007).

from cat (Joris and Yin, 1992), and the right panels [(B) and (D)] show corresponding model responses with matched carrier frequency (at CF), modulation frequency (100 Hz), and other stimulus conditions. Modulation depths were varied from 0 to 0.99, and each response is accompanied by a half-wave rectified version of the respective input AM stimuli (i.e., a modulation gain of 0 dB) to the right (two cycles of the responses are shown). Model responses are simulated for a stimulus level 17 dB above the threshold of the model fiber. For both physiological data and model predictions, the modulation of the response increases with modulation depth and appears more modulated than the corresponding half-wave rectified input stimulus in almost all cases. Because the offset adaptation of the model response shows a pause with a very slow recovery to spontaneous activity, the model AN fiber is less responsive in the dip of the envelope, and thus shows enhanced phase-locking, with responses clustered near the peak of the envelope.

The lower panels [(C) and (D)] of Fig. 12 show the synchronization coefficient[3] (R) and modulation gain [$20\log(2R/m)$, in decibels] derived from the corresponding histograms of the physiological data (A) and AN model responses (B) shown above. The dotted line shows the synchrony that would result if the response histogram perfectly followed the stimulus envelope. When the strength of synchrony for both model and physiological data is above the dotted line, the modulation gain is positive. Note that model responses show a higher synchronization coefficient than the corresponding data at higher modulation depths, which is due to the inclusion of fast power-law component in the model, as discussed further below. However, the previous AN model shows negative or near 0-dB gain (the model fiber in this case was substantially responsive in the dips of the AM stimulus and thus was not as well synchronized as the newer model responses).

Figure 13 illustrates the effects of modulation depth (m) and modulation frequency (f_m) on envelope synchrony as a function of AM stimulus level. The left panels [(A) and (C)] show physiological data from cat (Joris and Yin, 1992), and the corresponding model responses are shown in the right panels [(B) and (D)]. In this illustration of the effect of modulation depth [upper panels, (A) and (B)], the carrier frequency (set to the fiber's CF) was 2 kHz for a HSR fiber. The general non-monotonic shape of the synchrony-level function remains unchanged as the modulation depth is varied, but the range of levels over which significant synchrony is observed increases with increasing depth. The effect of modulation frequency on the synchrony-level function [lower panels, (C) and (D)] was studied for a HSR fiber with CF = 20 kHz and m=0.99. As for the physiological data, model synchrony-level functions superimpose at low f_m, although unlike the data, the curves are slightly separated at levels higher than the best modulation level (BML) (the level at which the response is maximum). At high f_m in both data and model, the entire synchrony-level curve shifts downward. In the PLA model responses, the BML remains almost constant as f_m increases (Fig. 13, lower panels), similar to that observed in cat (Joris and Yin, 1992) [but note that an upward shift in BML with increasing f_m was observed in guinea pig AN responses by Yates (1987)].

Physiological and model AN modulation transfer functions (MTFs) of high-CF fibers (>10 kHz) are shown in Fig. 14. Model MTFs were determined for a population of fibers with CFs spaced logarithmically (ranging from 10 to 20 kHz) at a level 10 dB above threshold for high, medium, and low SR fibers. Responses of 24 AN fibers (according to the proportions of SRs in the AN population) were simulated. Both physiological and model MTFs are low-pass in shape with cutoffs between ~600 and 1000 Hz. Each MTF is characterized by a shallow, slightly positive slope at f_m's below BMF

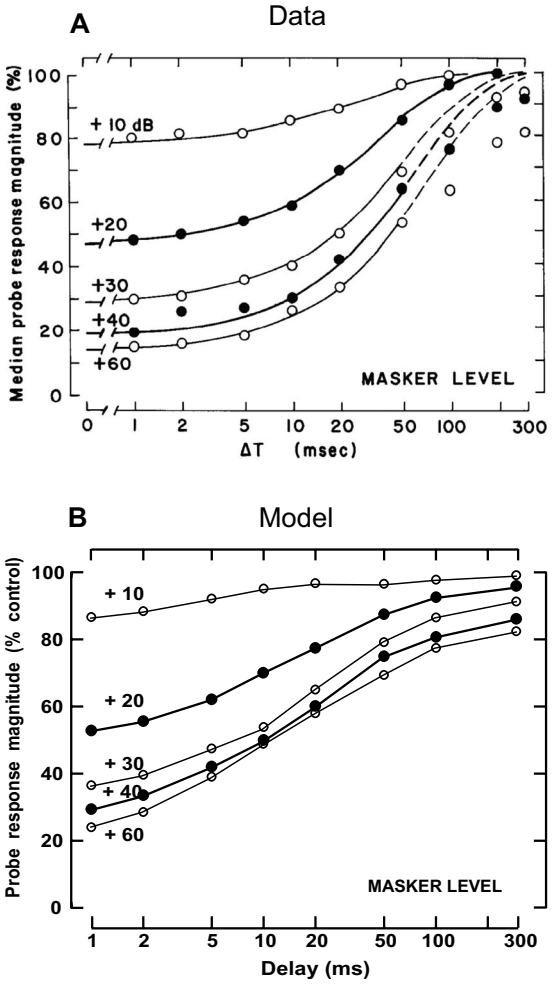


FIG. 11. Forward-masking recovery functions for a population of fibers; masker level is the parameter. Masker stimuli were tones with frequency matched to CF, 100-ms duration. Probe stimuli were also tones at CF, +20 dB, 15-ms duration. (A) Actual median recovery functions from 37 fibers with CFs ranging from 0.5 to 16 kHz. From Harris and Dallos (1979, with permission). (B) Model average recovery functions from ten CFs (HSR and LSR) spaced logarithmically (range of 0.5–16 kHz).

and by a sharp roll-off above BMF. Because the bandwidths of model AN fibers increase with CF, they are able to encode higher modulation frequencies; if this were the only factor limiting phase-locking to AM stimuli, the cut-off frequencies of the model MTFs would be expected to increase as a function of CF. However, as noted in Joris and Yin (1992), there exists an upper limit of f_m above which AN fibers cannot synchronize to the envelope because of low-pass filtering in the IHC, in addition to the progressive rejection of the sidebands by the sharp filtering in the cochlea, as discussed further below.

Figure 15 shows the relationship between AN fiber tuning-curve parameters (CF and bandwidth) and the MTF cut-off frequency. Left panels [(A) and (C)] show the physiological responses from cats (Joris and Yin, 1992), and the right panels [(B) and (D)] represent the corresponding model responses. Model responses were determined from a population of AN fibers (all operating at 10 dB above threshold) with CFs ranging from 250 Hz to 20 kHz (spaced logarithmically) for high (n=61), medium (n=23), and low (n=16) SR fibers. For better comparison to Joris and Yin (1992), medium SR fibers were included in the low SR group. Note that the abscissae in the lower panels differ because model responses had smaller bandwidths than the physiological data. This difference is explained by the fact that model responses were simulated using the 50th percentile of Q_{10} (CF/bandwidth) values (Zilany and Bruce, 2006) from Miller *et al.* (1997) which did not include the large range of bandwidths observed by Joris and Yin (1992). Because tuning bandwidth increases with CF, a positive correlation between the MTF cut-off frequency and CF is evident from both physiological data and model responses. However, MTF cut-off frequency saturates at higher CFs, which suggests that some mechanism in addition to peripheral bandpass filtering must exist to limit the response modulation. It is hypothesized here that the IHC low-pass filter is a candidate for this limitation, as discussed in detail in Sec. IV.

The effect of SR on maximum synchronization to f_m is shown in Fig. 16. The upper panel (A) shows physiological data from cats (Joris and Yin, 1992), and the lower panel (B) represents model responses. Model responses were determined from a population of AN fibers with CFs ranging from 250 Hz to 20 kHz, including high, medium, and low SR fibers. Stimuli were 10 dB above threshold for each model fiber, and the maximum synchrony was chosen from responses to a wide range of f_m's (10 Hz–2 kHz). Both physiological data and model responses show that low-CF fibers tend to have lower maximum synchrony than high-CF fibers with similar SRs. However, model responses do not show an inverse relationship between maximum synchrony and SR for high-CF fibers, in contrast to the physiological data. This discrepancy could be due to the fact that the parallel fast power-law component provides significant synchronized responses to the envelope, irrespective of the model fiber's SR.

F. Responses to noise stimuli

The shuffled autocorrelogram (SAC) and the cross-stimulus autocorrelogram (XAC) provide convenient and robust ways to quantify temporal information (discharge times) in response to wideband noise before and after polarity inversion (Joris, 2003; Louage *et al.*, 2004). SACs reveal that AN fibers are more temporally consistent (i.e., tend to discharge at the same point in time on repeated presentations of the same stimulus) in response to stochastic noise stimuli than in response to periodic tones. The normalized SAC also reveals how spikes are constrained in their timing jointly by cochlear filtering and phase-locking to fine-structure and envelope. The maximum SAC value, referred to as the central peak, is always reached at a delay near 0 ms. Joris (2003) argued that the central peak of the SAC reflects synchronization to different waveform features for fibers with different CFs. Responses of low- and high-CF fibers reflect phase-locking to fine-structure and envelope, respectively. The central peak also shows large differences across different classes of SRs. The shapes of the SAC and XAC change with increasing CF: for CFs above the range of pure-tone phase-locking, the SAC and XAC become indistinguishable.

The upper panels in Fig. 17 show the central-peak height of normalized SAC to broadband noise (70-dB SPL) as a

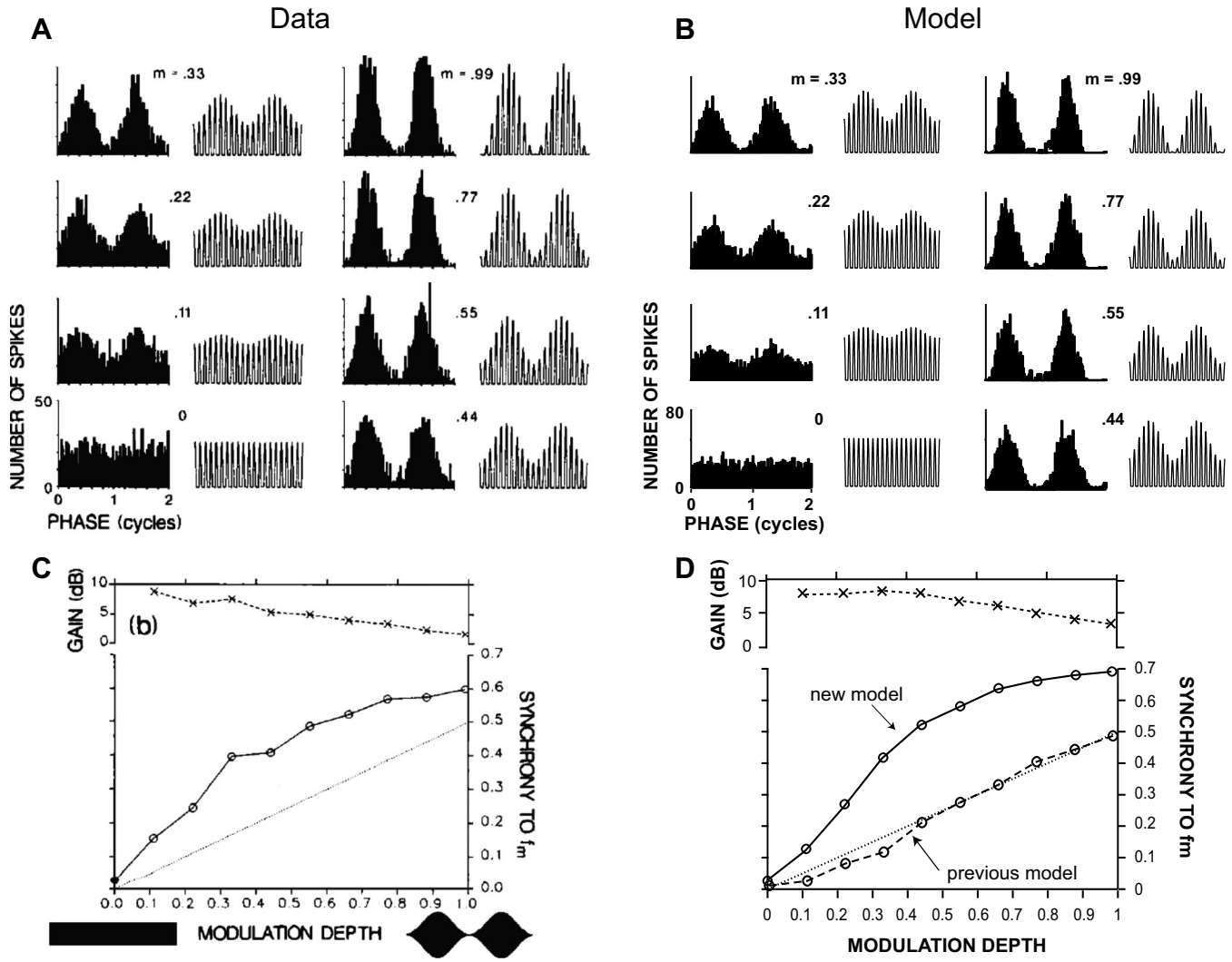


FIG. 12. Effect of increasing modulation depth (m) on synchrony for a HSR fiber with a CF of 20.2 kHz in response to amplitude-modulated tones with a carrier frequency matched to CF, modulation frequency (f_m)=100 Hz, and SPL = 49 dB (threshold = 32 dB SPL). Left panels [(A) and (C)] show the actual data and the right panels [(B) and (D)] represent the corresponding PLA model responses. Upper and lower panels show the period histograms, and their corresponding synchrony and gain, respectively, at different modulation depths. Modulation depths were varied from 0 to 0.99, and each histogram is flanked by a half-wave rectified version of the respective input AM stimulus to the right (two cycles of the responses are shown). [(A) and (C)] From Joris and Yin (1992, with permission). [(B) and (D)] Model period histograms and their corresponding synchrony and gain as a function of modulation depth using the same paradigm as employed in the experiment (the level of the stimulus is 17 dB above threshold). The dotted straight line represents 0-dB gain. Solid line with circles indicates the responses of the PLA model presented in this paper, and the dashed line with circles represents the responses of the previous model that had only exponential adaptation in the synapse model.

function of CF. Panel (A) shows the physiological responses from cats (Louage *et al.*, 2004), and the right panel (B) represents corresponding PLA model responses. Each point represents the response from a single fiber. Model responses were determined for a population of fibers with CFs ranging from 250 Hz to 20 kHz (20 fibers logarithmically spaced) for high (plus), medium (circle), and low (downward triangle) SR fibers. In both physiological data and PLA model responses, the height of the central peak decreases with CF but asymptotes for CFs near the limit of pure-tone phase-locking (4–5 kHz), where it sometimes barely exceeds unity (a value of 1 in the normalized SAC corresponds to no temporal correlation). For fibers of similar CF, there is a considerable range of peak heights in the physiological data. Interestingly, the SR distribution within that range is bimodal: generally low/medium-SR fibers have larger peak heights than high-SR fibers. This bimodality is not dependent on a particular choice of stimulus level and is also observed for responses obtained at a fixed suprathreshold level. Model responses are closest to the upper range of the peak heights in the physiological data, suggesting that discharge patterns in the model are more regular than in the data.

Figure 17(c) shows the ratio of XAC and SAC values at delay 0 for a population of AN fibers (Louage *et al.*, 2004), and Fig. 17(d) shows the corresponding PLA model responses. Model responses were determined for a population of fibers with CFs ranging from 250 Hz to 20 kHz (20 fibers logarithmically spaced) for high (plus), medium (circle), and low (downward triangle) SR fibers. As in the physiological data, the ratio of XAC and SAC values in the model responses has a sigmoidal relationship as a function of CF which illustrates a transition from the fine-structure coding at low CFs to envelope coding at high CFs (Louage *et al.*,

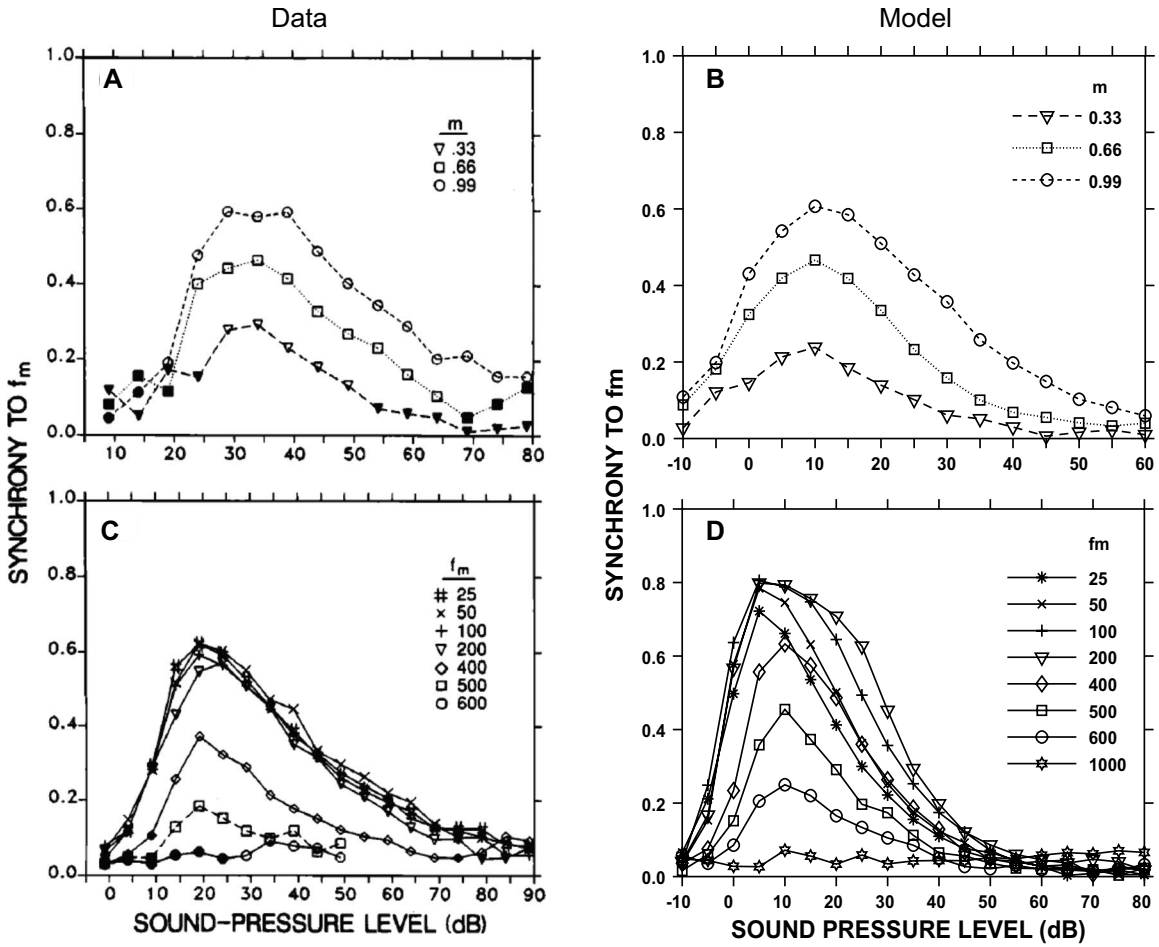


FIG. 13. Effect of modulation depth (m) and frequency (f_m) on the actual [left panels: (A) and (C)] and PLA model [right panels: (B) and (D)] synchrony-level function. In the study of the effect of modulation depth (upper panels), CF is at 2 kHz (HSR fiber). For the effect of modulation frequency (lower panels) on synchrony-level function, CF is at 20 kHz (HSR fiber) with m=0.99. [(A) and (C)] From Joris and Yin (1992, with permission). [(B) and (D)] Model responses using the same paradigm as in the experiment.

2004), and there is no apparent distinction across the SR groups.

IV. DISCUSSION

A. Achievement with regard to previous models

The PLA model is successful in describing a range of response properties of AN fibers that were not adequately addressed by previous models. Models having only exponential adaptation produce responses that do not scale with the duration of the stimulus, which affects the recovery after stimulus offset as well as long-term response properties of these models. Power-law dynamics significantly improved the offset-recovery response, which in turn provided better responses to forward-masking and AM signals. The model also successfully replicated the histogram of AN SRs using only three true SRs with long-term fluctuations. Due to the addition of fGn to the input of the slow power-law adaptation path, the model responses are positively correlated over the long term and are negatively correlated over the short term (result not shown).

The IHC-AN synapse model presented in this study has two parallel paths with slow and fast power-law dynamics, following a stage with exponential adaptation. The model was thus capable of replicating additivity seen in AN responses to stimulus increment paradigm. Both the slow and fast power-law adaptation components contributed to higher synchronized responses to the envelope of AM signals and also to pure tones at low frequencies. It is worth mentioning that this model was also capable of producing strongly synchronized responses of high-CF fibers to low-frequency tones at high stimulus levels (Joris *et al.*, 1994); for example, the synchronization coefficient of a model AN fiber with CF = 10 kHz to a 80 or 90 dB SPL, 800 Hz tone is ~0.9, whereas the maximum synchronization coefficient of an 800-Hz model fiber to a tone at CF is ~0.83 (Johnson, 1980).

One of the important achievements of the PLA model is that it can explain two seemingly contradictory aspects of forward-masking data reported by Harris and Dallos (1979) and Young and Sachs (1973). Harris and Dallos (1979) showed that the reduction in probe responses saturated at higher levels as the masker-evoked responses saturate at higher levels. However, Young and Sachs (1973) showed that the time course of recovery continues to increase with masker/exposure level, even though the masker discharge rates saturate at higher levels. The duration of the forward masker (100 ms) in Harris and Dallos (1979) is much shorter

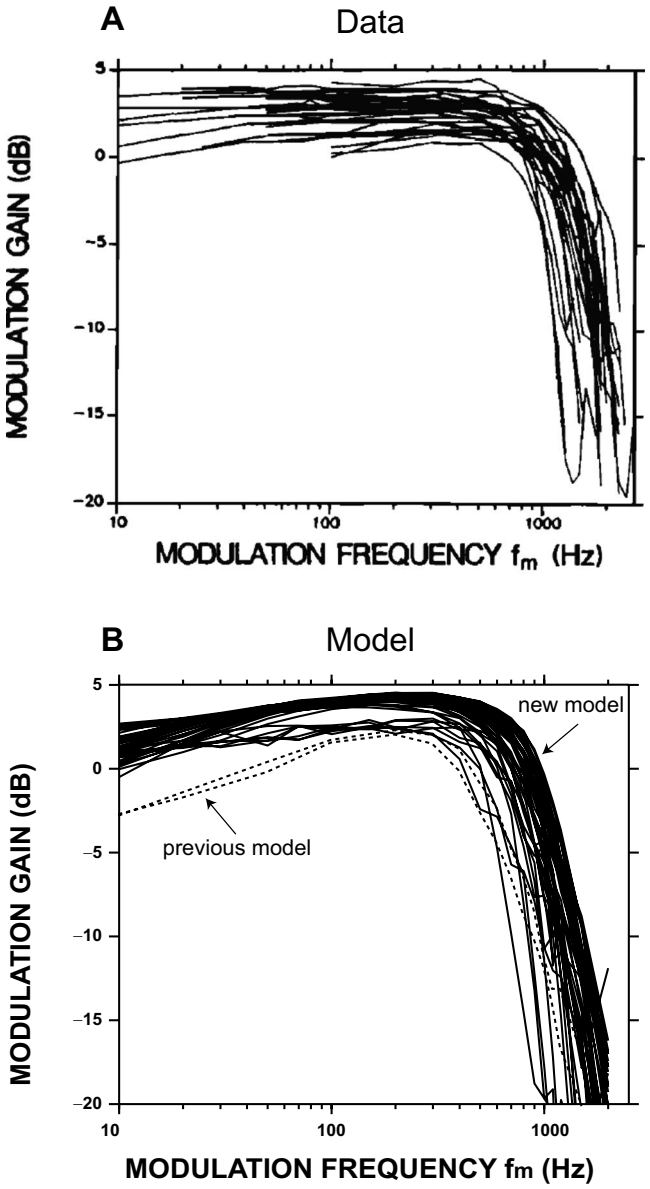


FIG. 14. MTFs of high CF (>10 kHz) fibers. Upper panel shows the actual MTFs from cat, and the lower panel represents PLA model responses. (A) From Joris and Yin (1992, with permission). (B) Model MTFs for a population of fibers with CFs spaced logarithmically (range of 10–20 kHz) at level 10 dB above threshold for high, medium, and low SR fibers. Responses for 24 AN fibers (according to the proportions of the distribution of SRs) are simulated. Solid lines show the responses of the PLA model presented in this paper, and the two dashed lines (CF at 10 and 20 kHz) indicate the responses of the previous model.

than that of Young and Sachs (1973), which was 60 s. In response to a short duration masker and at small masker-probe delays, the fast power-law component of the PLA model remains almost shut off and contributes very little to the probe response at higher masker levels; it is the slow power-law adaptation component that contributes to the probe response at these higher levels. As the input to the slow power-law component (i.e., exponential output) saturates at higher levels, the reduction in probe response also nearly saturates in the output of the slow power-law adaptation. Therefore, the reduction in probe response at short masker-probe delays becomes nearly saturated at higher levels, as Harris and Dallos (1979) observed. This is also consistent with another observation by Smith (1977) that for shorter forward maskers, recovery depends on the discharge rate in response to the masker rather than on absolute masker intensity. However, when the duration of the masker and masker-probe delay is sufficiently long, both fast and slow power-law components contribute to the recovery and exhibit recovery time courses that increase with level, as explained in Sec. III B 2.

B. Source of power-law adaptation

Although power-law dynamics has been prevalent in descriptions of sensory adaptation, identification of their physical basis remains enigmatic. In some preparations, the site of power-law adaptation has been located in the conversion of the receptor potential into action potentials (French and Torkkeli, 2008). French (1984) observed no detectable adaptation in the receptor potential in cockroach tactile spine, whereas power-law adaptation exists in the action potential trains of the associated somatosensory neurons (Chapman and Smith, 1963). Even direct electrical stimulation of action potentials, which bypassed the mechanotransduction stage, produced the same power-law adaptation (French, 1984), suggesting that post-synaptic membrane dynamics could be responsible for the observed adaptation. In the visual system, studies of temporal contrast in mammalian (rabbit and guinea pig) retina by Smirnakis *et al.* (1997) showed that the timescale of adaptation varies as a function of the period between stimulus switches, indicating the presence of multiple timescales or power-law adaptation.

Recently, Zhang *et al.* (2007) observed spike-rate adaptation in AN fiber responses to stimulation by a cochlear implant using high-rate pulse trains (of 300-ms duration), which suggests that adaptation is not purely a synaptic phenomenon. They fitted the rate vs time functions (adaptation at the onset) with two-exponent models and reported time-constants (rapid 8 ms, and short-term 80 ms) which were slightly higher than those of similar acoustic studies. Although these time constants have little dependence on onset spike rate, they do show a strong relationship with input stimulus pulse rate. On the other hand, in simultaneous recordings from IHCs and AN fiber terminals, Goutman and Glowatzki (2007) observed that during a 1-s IHC depolarization, the synaptic response was depressed more than 90%, indicating that synaptic depression was the main source for adaptation in the AN. In their experimental data, the time course of transmitter release was fitted with three exponential transient components (with time constants of ~2, ~18, and ~176 ms) in addition to a longer-term component that they described as being "robust" to adaptation. However, as the duration of their measurements was relatively short, it is not clear whether the adaptation in the release would scale with the duration of the stimulus (which would suggest the presence of power-law dynamics of adaptation).

In the above experiments, only responses at the onset were investigated. However, there exists a substantial body of experimental data describing adaptation to various acoustic stimulus features, such as responses to stimulus offset, forward masking, and increment/decrement paradigms. Re-

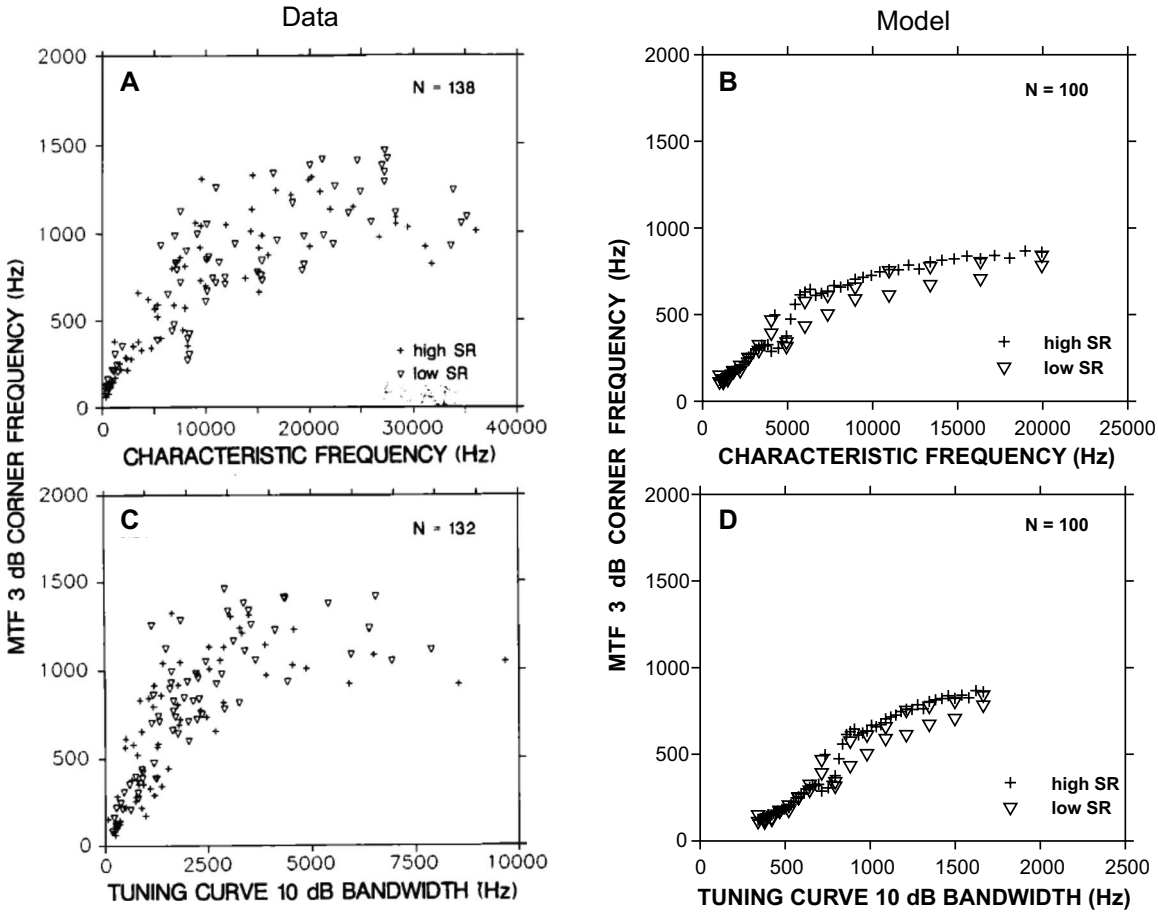


FIG. 15. MTF-3-dB cut-off frequencies vs CF and 10-dB bandwidth for high (plus) and low (down triangle) SR fibers. Left panels [(A) and (C)] show the actual responses from cat, and the right panels [(B) and (D)] represent the PLA model responses. (A) and (C) From Joris and Yin (1992, with permission). [(B) and (D)] Model responses for a population of fibers with CFs ranging from 250 Hz to 20 kHz (100 fibers spaced logarithmically) for high (61), medium (23), and low (16) SR fibers. Medium SR fibers are included in the low SR fibers, as treated in Joris and Yin (1992). Notice that the abscissae in the right panels (model responses) are different from those in the left panels (actual responses).

sponses to similar stimulus paradigm are required in the above-mentioned experiments to elucidate the degree of contribution by synaptic and membrane mechanisms to the adaptation observed with acoustic excitation.

The strength of onset adaptation to acoustic stimuli seems more consistent across AN fibers, whereas the strength of suppression at offset seems to vary across fibers (even with similar SRs) (Kiang, 1965; Harris and Dallos, 1979). Similarly, Zhang *et al.* (2007) observed that some fibers were strong adapters and others showed weak adaptation in their electrical stimulation experiment, indicating that membrane dynamics might be responsible for the variable adaptation seen in the offset and long-term response properties of the AN. Also, in general, neural dynamics are more likely to give rise to power-law rather than exponential adaptation.

C. Factors influencing the MTF

MTFs at the level of the AN are characterized by low-pass filter shapes with sharp roll-offs and positive gains (ranging 0–5 dB) in the low-pass region. The offset adaptation properties of the IHC-AN synapse account for enhanced phase-locking to the stimulus envelope in AN fibers. As mentioned earlier, both the slow and fast power-law adaptation components of the model contributed to this synchronization and resulted in positive modulation gains in the model MTFs. However, the cut-off frequency (i.e., bandwidth) and the slope of the roll-off in model MTFs are slightly different than those of physiological MTFs (Fig. 14). At least two filtering actions by different mechanisms limit the frequency above which the AN fiber's instantaneous discharge rate is no longer modulated at f_m: mechanical and temporal filtering (Greenwood and Joris, 1996). The local basilar partition motion driving the IHC is a mechanically bandpass-filtered version of the cochlear input. The Q_{10} value, specified as a function of CF, sets the bandwidth of this filter in the model. Placing the carrier frequency at fiber CF, this filter progressively removes the sideband components of the AM stimulus in the local motion as f_m increases. The removal of sideband components effectively reduces the envelope amplitude variation and thus influences the MTF cut-off frequency. In model responses, higher $Q10$ values (i.e., lower bandwidths) at a particular CF produce MTFs with lower cut-off frequencies (results not shown). It is to be noted that the model Q_{10} values are significantly higher at higher CFs than those in Joris and Yin (1992), and hence the cut-off frequencies of the model MTFs are lower.

The temporal filter resides in the stage between mechanical motion and AN spikes and acts as a low-pass filter

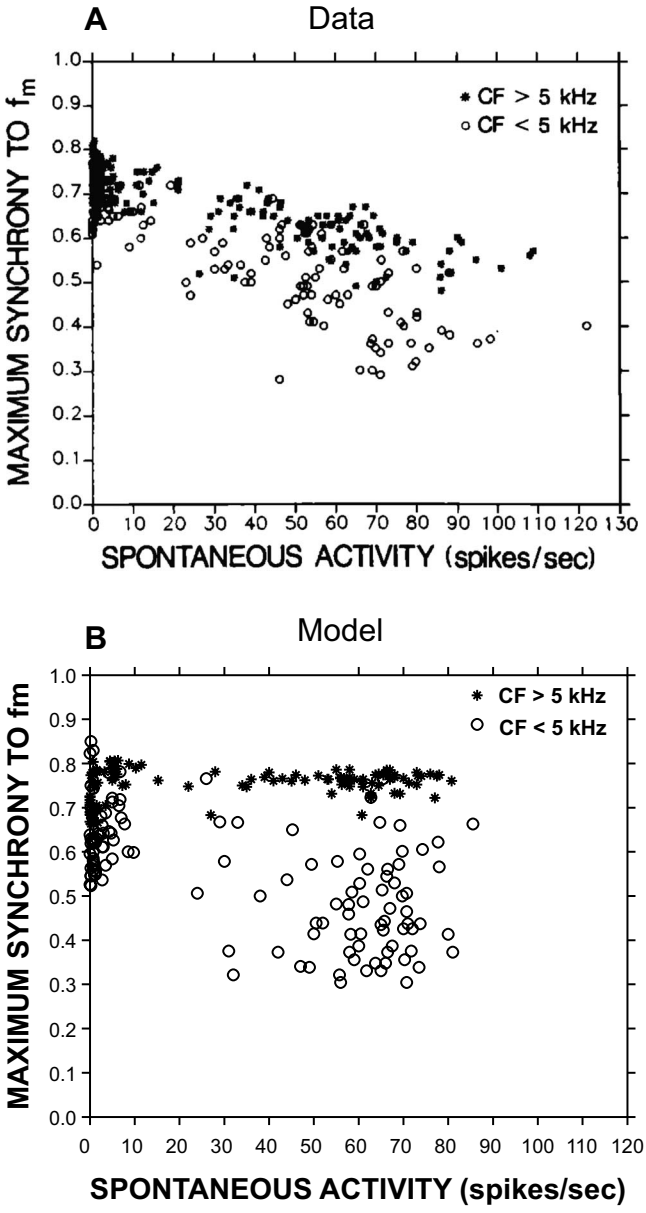


FIG. 16. Effect of SR on maximum synchronization to f_m. Upper panel shows the actual data from cat, and the lower panel shows the PLA model responses. (A) From Joris and Yin (1992, with permission). (B) Model responses for a population of fibers with CFs ranging from 250 Hz to 20 kHz for high, medium, and low SR fibers. Each fiber operates at 10 dB above threshold, and the maximum synchrony is chosen from responses to a wide range of f_m (10 Hz–2 kHz).

that limits synchronization of AN responses to temporal variations in the IHC input. This added constraint on the bandpass-filtered signal further changes the magnitude of synchronization to envelope. A seventh-order low-pass filter with a cut-off frequency of 3.0 kHz was used in the model to represent this stage. However, both the order and the cut-off frequency of this low-pass filter influence the MTF shape (results not shown). A higher-order filter results in a MTF with a sharper roll-off, and a higher cut-off frequency of this temporal filter causes a higher cut-off frequency in the model MTF, unless it is already limited by the bandwidth of the basilar membrane (mechanical) filter (i.e., Q_{10} value). Therefore, it is possible to accurately replicate individual physiological MTFs using appropriate model Q_{10} values and the correct order and cut-off frequency of the IHC low-pass filter. This result illustrates that accurate modeling can identify or predict potential mechanisms of certain processes where direct physiological study is either very cumbersome or impossible.

D. Implementing SRs: Three rather than two true SRs

Jackson and Carney (2005) showed that a model with only two or three SRs with long-term fluctuations could describe the histogram of AN SRs in cat. In the case of two true SRs, instead of using an inhomogeneous Poisson process, they employed a Poisson-equivalent integrate-and-fire model in which negative values of the driving function (not rectified) have a negative effect on the output. In particular, the negative input values reduce the value of the running integral that accumulates toward threshold and thus delay the time of discharge occurrence. Although this property achieves the distribution of low SRs in the histogram, it produces AN responses that are inconsistent with physiological observations. For instance, in the PSTHs of low SR fibers in response to tones, the peak onset response strongly depends on the silent interval between stimulus offset and the next onset; shorter intervals reduce the onset responses because the fiber does not have enough time to recover, and on the other hand, sufficiently longer silent intervals produce sharp, large-magnitude peaks at the onset. However, in Jackson and Carney's (2005) model with two true SRs, the negative input values for low SR fibers (which tend to have driving functions with more negative values) would result in the opposite pattern of response: longer silent intervals will accumulate more negative values, which will then result in greater reduction in the onset than for short intervals. In the three true SR model, negative inputs do not contribute to the running integral, and thus this unwanted result is not observed. This result suggests that the three true SR model better accounts for the observed AN responses as well as for the distribution of SRs. In the PLA model, the discharge generator (inhomogeneous Poisson process) section was implemented in such a way that the negative driving function has no effect on the output responses (i.e., equivalent to rectification of the driving signal). Three fGn parameter sets designed corresponding to three true SRs were able to describe the SR histogram of AN fibers, while maintaining other features of AN responses to a wide variety of stimuli.

Although fGn with appropriate parameters was added in the slow power-law adaptation path, the physiological correlate of this noise along the auditory-periphery is not clear. Kelly *et al.* (1996) reported that this noise is independent of CF and SR of the AN fiber. They argued that this fractal phenomenon originates either in the IHC or at the synaptic junction between IHC and AN fibers. Also, Teich and Lowen (1994) speculated on a number of possible origins of the observed fractal behavior, such as the slow decay of intracellular calcium in the hair-cell receptor or fractal ion-channel statistics.

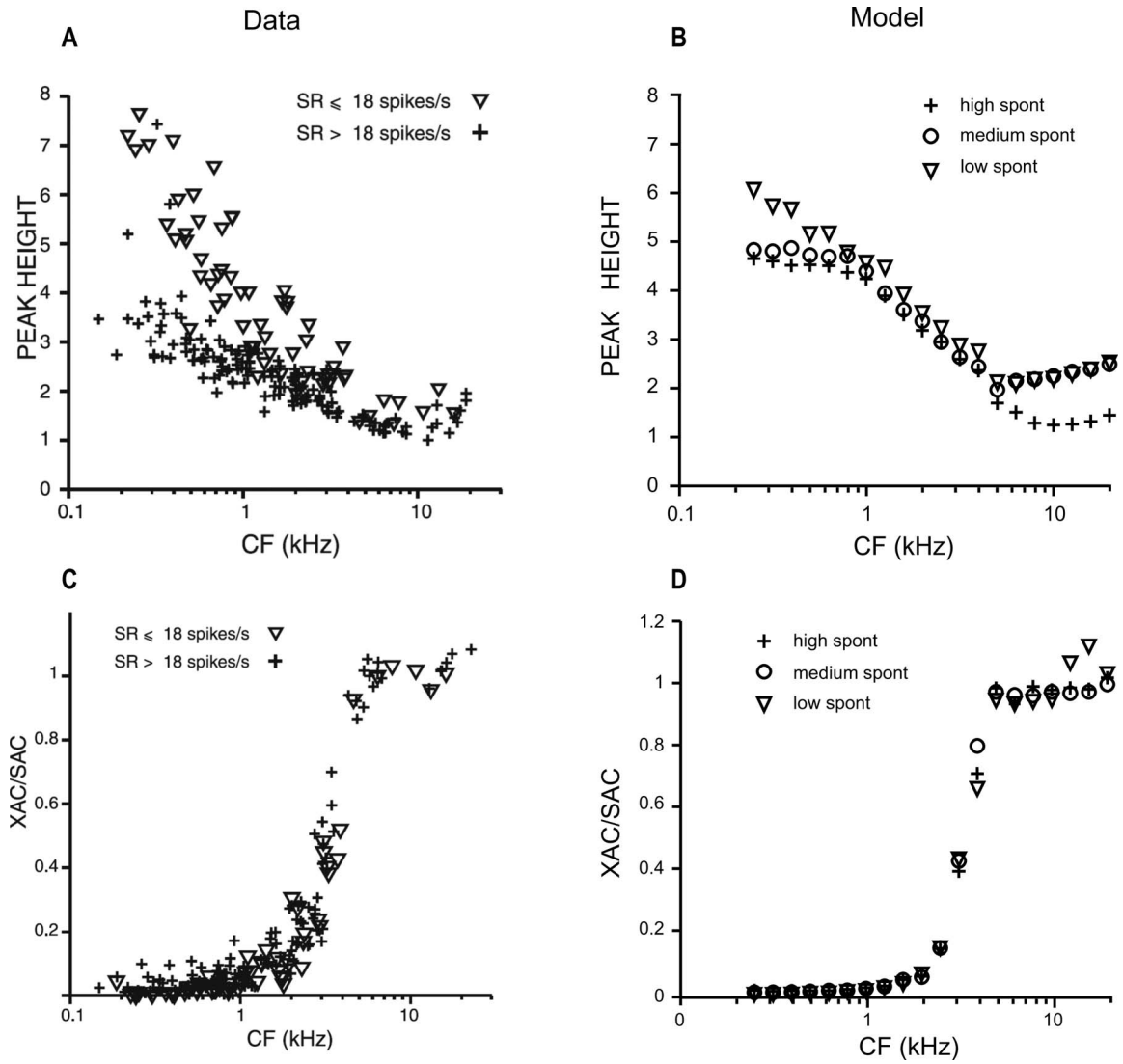


FIG. 17. Upper panels: central-peak height of normalized SAC to broadband noise (70-dB SPL) vs CF for a population of AN fibers. Lower panels: ratio of the value at delay 0 of XAC and SAC for a population of fibers. Left panels [(A) and (C)] show the actual responses from cat, and the right panels [(B) and (D)] represent the corresponding PLA model responses. Each point represents response from a single fiber. [(A) and (C)] From Louage *et al.* (2004, with permission). [(B) and (D)] Model responses for a population of fibers with CFs ranging from 250 Hz to 20 kHz (20 fibers logarithmically spaced) for high (plus), medium (circle), and low (downward triangle) SR fibers.

E. Implications for complex sounds and psychophysics

In general, adaptation yields an efficient sensory code by removing redundant information inherent in the environmental cues. The natural acoustic environment is made up mostly of transients rather than constant stimuli. Adaptation helps to efficiently encode stimuli with statistics that vary in time (Delgutte, 1980). To encode efficiently, a neural system must change its coding strategy as the distribution of stimuli changes. Power-law dynamics, possessing no privileged timescales, is invariant with respect to changes in temporal scale, and such a system could therefore adjust its effective adaptation timescale to the environment. Recently, studies in the auditory midbrain (Dean *et al.*, 2005) and cortex (Watkins and Barbour, 2008) show that neurons respond to recent stimulus history by adapting their response functions according to the statistics of the stimulus, alleviating the so-called "dynamic range problem." However, the mechanism and origin of this adaptation along the auditory pathway remain unclear. An auditory-nerve model with appropriate long-term dynamics (power-law-like) in the IHC-AN synapse, such as that presented in this study, could successfully account for this adaptation, including the time course of adaptation. Further studies with this model will pursue this phenomenon.

Many psychophysical studies have mapped out the magnitude and time course of forward masking using a variety of stimulus paradigms (Hanna *et al.*, 1982; Zwicker, 1984; Dau *et al.*, 1996b). Several fundamental features of these data cannot be easily explained with the responses of single AN fibers (Relkin and Turner, 1988). Sub-cortical neural processing appears to have strong influence on perception in these tasks (Nelson *et al.*, 2009), but specific mechanisms underlying the transformation of forward-masked stimuli have not been carefully tested with experiments or models. The phe-

nomenological model described here provides a realistic front-end to test central models with an input that reasonably predicts several related sets of AN data.

Recently, Dau *et al.* (1996a) developed a model of signal processing in the auditory system to explain the psychophysical thresholds for various masking conditions (Dau *et al.*, 1996b). They employed an adaptation stage in the peripheral system that has five feedback loops, connected in series, with five different time constants. In each loop, the output is the input signal divided by a low-pass-filtered version of the output, similar to a single-loop model proposed by Siebert and Gambardella (1968) to account for the effects of stimulus level and duration on adaptation in the discharge rates of AN fibers. Although these models can address rate adaptation to some extent, they do not have power-law dynamics because the time constants of the low-pass filters are fixed. The model by Dau *et al.* (1996a) explained the psychophysical data well, except for the forward-masked thresholds obtained with brief maskers, which were too high compared to the measured data. They pointed out that it was the adaptation stage in the model that was responsible for this behavior. As the time constants (ranging from 5 to 500 ms) of the low-pass filters in the adaptation-loop model are fixed irrespective of the masker duration, recovery in the masker offset does not scale appropriately with the duration of the masker. Thus, although the model explained the forward-masked thresholds for long maskers, it failed to address the thresholds for brief maskers. In this regard, the new AN model with power-law dynamics in the adaptation stage would be a better candidate to explain these monaural psychophysical data as well as other binaural masking data (Breebaart *et al.*, 2001) that also employed the peripheral model of Dau *et al.*(1996a).

One of the most obvious features of a speech signal is amplitude modulation, and much of the information of speech appears to be carried in these changes rather than in the relatively stationary aspects of speech (Shannon *et al.*, 1995). Recent psychophysical models of AM perception assume that a population of modulation-selective filters provides information about a signal's temporal envelope to higher processing centers (e.g., Dau *et al.*, 1997; Ewert *et al.*, 2002). As the PLA model can reliably produce the MTFs of AN fibers, the output of this model can be used as front-end to models for higher auditory centers to test realistic neural-encoding hypotheses that may be used by the auditory system to encode envelope modulations.

F. Limitations

Despite its success in explaining a number of AN response features, there are a number of limitations in the PLA model that require further study. It was assumed that there is no adaptation in the voltage responses of the IHC, but recent studies suggest that there is indeed some adaptation at this level (Kros and Crawford, 1990; Zeddies and Siegel, 2004; Jia *et al.*, 2007; Beurg *et al.*, 2008). It would be important to explore the contribution of IHC adaptation to AN responses, especially at the onset and offset of tone bursts and in response to AM stimuli.

The PLA model does not capture the relationship between maximum synchrony to AM stimuli and SR, particularly for high-SR fibers. Physiological data show an inverse relationship between these metrics, whereas the model responses are nearly constant as a function of SR at a high value of synchrony. As mentioned in Sec. III, the fast power-law adaptation component of the model yields highly synchronized responses to AM signals irrespective of SR, which explains the high maximum synchrony to modulation frequency. The ability of the model to relate SRs to different response properties is thus limited, and further exploration is needed in this regard.

The actual power-law adaptation is computationally very expensive. Although an approximation to the power-law was implemented by an IIR filter, the actual implementation was required to replicate the very long-term response properties (Fig. 6) with this model.

Although the PLA model captures a wide range of AN response properties, physiological correlates of the model architecture are not evident from existing studies. More experimental data are needed to build a more biophysically based model or to justify the proposed phenomenological model.

V. CONCLUSION

This paper presents a phenomenological model of the auditory periphery with a new IHC-AN synapse model that has adaptation at different time scales. Several important adaptation measures other than the onset response, such as recovery after offset (Harris and Dallos, 1979), responses to increments and decrements (Smith *et al.*, 1985), and conservation (Westerman and Smith, 1987), were satisfactorily captured by this model. The PLA model is thus capable of accurately predicting several sets of AN data such as the amplitude-modulation transfer function and forward masking, which the exponential adaptation model clearly fails to address. The success of the power-law adaptation in describing a wide range of AN responses indicates a possible mechanism of adaptation, other than the classically described exponential adaptation, in the IHC-AN synapse and/or in the AN membrane.

ACKNOWLEDGMENTS

We appreciate helpful discussions and communications during the development of this model with Dr. Robert Smith and Dr. Philip Joris. The authors also thank Dr. Michael Heinz for comments on an earlier version of the manuscript. The suggestions of two anonymous reviewers were invaluable in improving this manuscript. This research was supported by NIH-NIDCD R01-01641 (M.S.A.Z., L.H.C.) and CIHR Grant Nos. 54023 (I.C.B.) and F32-009164 (P.C.N.).

[1]The time required for the PLA model with the actual power-law implementation to simulate ten repetitions of a 1-s duration stimulus (i.e., a total duration of 10 s) is ~50 times greater than the time taken by the previous model (Zilany and Bruce, 2007). However, the computational times for the previous model and for the PLA model with the approximate power-law implementation are nearly the same.

[2]Both approximate and actual implementations are available in the code.

[3]Synchronization coefficient, or vector strength, (R) is a dimensionless

measure of phase-locking and is defined for a particular frequency as the ratio of the magnitude of the synchronized response at that frequency and the average response rate of the fiber (Johnson, 1980).

Abbas, P. J. (**1979**). "Effects of stimulus frequency on adaptation in auditory-nerve fibers," J. Acoust. Soc. Am. **65**, 162–165.
Beurg, M., Nam, J.-H., Crawford, A., and Fettiplace, R. (**2008**). "The actions of calcium on hair bundle mechanics in mammalian cochlear hair cells," Biophys. J. **94**, 2639–2653.
Breebart, J., van de Par, S., and Kohlrausch, A. (**2001**). "Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters," J. Acoust. Soc. Am. **110**, 1105–1117.
Brown, M. C., and Stein, R. B. (**1966**). "Quantitative studies on the slowly adapting stretch receptor of the crayfish," Kybernetik **3**, 175–185.
Bruce, I. C., Sachs, M. B., and Young, E. D. (**2003**). "An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses," J. Acoust. Soc. Am. **113**, 369–388.
Camera, G. L., Rauch, A., Thurbon, D., Luscher, H. R., Senn, W., and Fusi, S. (**2006**). "Multiple time scales of temporal response in pyramidal and fast spiking cortical neurons," J. Neurophysiol. **96**, 3448–3464.
Carney, L. H. (**1993**). "A model for the responses of low-frequency auditory-nerve fibers in cat," J. Acoust. Soc. Am. **93**, 401–417.
Chapman, K. M., and Smith, R. S. (**1963**). "A linear transfer function underlying impulse frequency modulation in a cockroach mechanoreceptor," Nature (London) **197**, 699–700.
Cooke, M. P. (**1986**). "A computer model of peripheral auditory processing," Speech Commun. **5**, 261–281.
Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2892–2905.
Dau, T., Püschel, D., and Kohlrausch, A. (**1996a**). "A quantitative model of the effective signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622.
Dau, T., Püschel, D., and Kohlrausch, A. (**1996b**). "A quantitative model of the effective signal processing in the auditory system. II. Simulations and measurements," J. Acoust. Soc. Am. **99**, 3623–3631.
Dean, I., Harper, N. S., and McAlpine, D. (**2005**). "Neural population coding of sound level adapts to stimulus statistics," Nat. Neurosci. **8**, 1684–1689.
Delgutte, B. (**1980**). "Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers," J. Acoust. Soc. Am. **68**, 843–857.
Drew, P. J., and Abbott, L. F. (**2006**). "Models and properties of power-law adaptation in neural system," J. Neurophysiol. **96**, 826–833.
Ewert, S. D., Verhey, J. L., and Dau, T. (**2002**). "Spectro-temporal processing in the envelope-frequency domain," J. Acoust. Soc. Am. **112**, 2921–2931.
Fairhall, A. L., Lewen, G. D., Bialek, W., and de Ruyter van Steveninck, R. R. (**2001**). "Efficiency and ambiguity in an adaptive neural code," Nature (London) **412**, 787–792.
French, A. S. (**1984**). "The receptor potential and adaptation in the cockroach tactile spine," J. Neurosci. **4**, 2063–2068.
French, A. S., and Torkkeli, P. H. (**2008**). "The power law of sensory adaptation: Simulation by a model of excitability in spider mechanoreceptor neurons," Ann. Biomed. Eng. **36**, 153–161.
Furukawa, T., and Matsuura, S. (**1978**). "Adaptive rundown of excitatory postsynaptic potentials at synapses between hair cells and eight nerve fibers in the goldfish," J. Physiol. (London) **276**, 193–209.
Furukawa, T., Mikuki, K., and Matsuura, S. (**1982**). "Quantal analysis of a decremental response at hair cell-afferent fiber synapse in the goldfish sacculus," J. Physiol. (London) **322**, 181–195.
Goutman, J. D., and Glowatzki, E. (**2007**). "Time course and calcium dependence of transmitter release at a single ribbon synapse," Proc. Natl. Acad. Sci. U.S.A. **104**, 16341–16346.
Greenwood, D. D., and Joris, P. X. (**1996**). "Mechanical and temporal filtering as codeterminants of the response by cat primary fibers to the amplitude-modulated signal," J. Acoust. Soc. Am. **99**, 1029–1039.
Hanna, T. E., Robinson, D. E., Shiffrin, R. M., and Gilkey, R. H. (**1982**). "Forward masking of diotic and dichotic clicks by noise," J. Acoust. Soc. Am. **72**, 1171–1177.
Harris, D. M., and Dallos, P. (**1979**). "Forward masking of auditory nerve fiber responses," J. Neurophysiol. **42**, 1083–1107.
Hewitt, M. J., and Meddis, R. (**1991**). "An evaluation of eight computer models of mammalian inner hair-cell function," J. Acoust. Soc. Am. **90**, 904–917.
Jackson, B. S. (**2003**). "Consequences of long-range temporal dependence in neural spike trains for theories of coding and processing," Ph.D. thesis, Syracuse University, Syracuse, NY (**2003**).
Jackson, B. S., and Carney, L. H. (**2005**). "The spontaneous-rate histogram of the auditory nerve can be explained by only two or three spontaneous rates and long-range dependence," J. Assoc. Res. Otolaryngol. **6**, 148–159.
Jia, S., Dallos, P., and He, D. Z. (**2007**). "Mechanoelectric transduction of adult inner hair cells," J. Neurosci. **27**, 1006–1014.
Johnson, D. (**1980**). "The relationship between spike rate and synchrony in responses to auditory-nerve fibers to single tones," J. Acoust. Soc. Am. **68**, 1115–1122.
Joris, P. X. (**2003**). "Interaural time sensitivity dominated by cochlea-induced envelope patterns," J. Neurosci. **23**, 6345–6350.
Joris, P. X., Carney, L. H., Smith, P. H., and Yin, T. C. T. (**1994**). "Enhancement of neural synchronization in the anteroventral cochlear nucleus I: Responses to tones at the characteristic frequency," J. Neurophysiol. **71**, 1022–1036.
Joris, P. X., and Yin, T. C. T. (**1992**). "Responses to amplitude-modulated tones in the auditory nerve of the cat," J. Acoust. Soc. Am. **91**, 215–232.
Kelly, O. E., Johnson, D. H., Delgutte, B., and Cariani, P. (**1996**). "Fractal noise strength in auditory-nerve fiber recordings," J. Acoust. Soc. Am. **99**, 2210–2220.
Kiang, N. Y.-S. (**1965**). "Discharge patterns of single fibers in the cat's auditory nerve," M.I.T. Research Monograph No. 35 (Technology Press, Boston, MA).
Kiang, N. Y.-S. (**1990**). "Curious oddments of auditory-nerve studies," Hear. Res. **49**, 1–16.
Kros, C. J., and Crawford, A. C. (**1990**). "Potassium currents in inner hair cells isolated from the guinea-pig cochlea," J. Physiol. (London) **421**, 263–291.
Leopold, D. A., Murayama, Y., and Logothetis, N. (**2003**). "Very slow activity fluctuations in monkey visual cortex: Implications for functional brain imaging," Cereb. Cortex **13**, 422–433.
Liberman, M. C. (**1978**). "Auditory-nerve response from cats raised in a low-noise chamber," J. Acoust. Soc. Am. **63**, 442–455.
Louage, D. H. G., Heijden, M. v. d., and Joris, P. X. (**2004**). "Temporal properties of responses to broadband noise in the auditory nerve," J. Neurophysiol. **91**, 2051–2065.
Lundstrom, B. N., Higgs, M. H., Spain, W. J., and Fairhall, A. L. (**2008**). "Fractional differentiation by neocortical pyramidal neurons," Nat. Neurosci. **11**, 1335–1342.
Meddis, R. (**1986**). "Simulation of mechanical to neural transduction in the auditory receptor," J. Acoust. Soc. Am. **79**, 702–711.
Meddis, R. (**1988**). "Simulation of auditory-neural transduction: Further studies," J. Acoust. Soc. Am. **83**, 1056–1063.
Meddis, R., and O'Mard, L. P. (**2005**). "A computer model of the auditory nerve response to forward masking stimuli," J. Acoust. Soc. Am. **117**, 3787–3798.
Miller, R. L., Schilling, J. R., Franck, K. R., and Young, E. D. (**1997**). "Effects of acoustic trauma on the representation of the vowel /ɛ/ in cat auditory nerve fibers," J. Acoust. Soc. Am. **101**, 3602–3616.
Moser, T., and Beutner, D. (**2000**). "Kinetics of exocytosis and endocytosis at the cochlear inner hair cell afferent synapse of the mouse," Proc. Natl. Acad. Sci. U.S.A. **97**, 883–888.
Müller, M., and Robertson, D. (**1991**). "Relationship between tone burst discharge pattern and spontaneous firing rate of auditory nerve fibers in the guinea-pig," Hear. Res. **57**, 63–70.
Nelson, P. C., and Carney, L. H. (**2004**). "A phenomenological model of peripheral and central neural responses to amplitude-modulated tones," J. Acoust. Soc. Am. **116**, 2173–2186.
Nelson, P. C., Smith, Z. M., and Young, E. D. (**2009**). "Wide dynamic range forward suppression in marmoset inferior colliculus neurons is generated centrally and accounts for perceptual masking," J. Neurosci. **29**, 2553–2562.
Oono, Y., and Sujaku, Y. (**1974**). "A probabilistic model for discharge patterns of auditory nerve fibers," Trans. Inst. Elect. Comm. Eng. (Japan) **57**, 35–36.
Oono, Y., and Sujaku, Y. (**1975**). "A model for automatic gain control observed in the firing of primary auditory neurons," Trans. Inst. Elect. Comm. Eng. **58**, 61–62.
Payton, K. L. (**1988**). "Vowel processing by a model of the auditory periphery: A comparison to eighth-nerve responses," J. Acoust. Soc. Am. **83**, 145–162.
Raman, I. M., Zhang, S., and Trussell, L. O. (**1994**). "Pathway-specific

variants of AMPA receptors and their contribution to neuronal signaling," J. Neurosci. **14**(18), 4998–5010.
Relkin, E. M., and Doucet, J. R. (**1991**). "Recovery from prior stimulation. I: Relationship to spontaneous firing rates of primary auditory neurons," Hear. Res. **55**, 215–222.
Relkin, E. M., and Turner, C. W. (**1988**). "A reexamination of forward masking in the auditory nerve," J. Acoust. Soc. Am. **84**, 584–591.
Rhode, W. S., and Smith, P. H. (**1985**). "Characteristics of tone-pip response patterns in relationship to spontaneous rate in cat auditory nerve fibers," Hear. Res. **18**, 159–168.
Ross, S. (**1982**). "A model of the hair cell-primary fiber complex," J. Acoust. Soc. Am. **71**, 926–941.
Ross, S. (**1996**). "A functional model of the hair cell-primary fiber complex," J. Acoust. Soc. Am. **99**, 2221–2238.
Schnee, M. E., Lawton, D. M., Furness, D. N., Benke, T. A., and Ricci, A. J. (**2005**). "Auditory hair cell-afferent fiber synapses are specialized to operate at their best frequencies," Neuron **47**, 243–254.
Schroeder, M. R., and Hall, J. L. (**1974**). "Model for mechanical to neural transduction in the auditory receptor," J. Acoust. Soc. Am. **55**, 1055–1060.
Schwid, H. A., and Geisler, C. D. (**1982**). "Multiple reservoir model of neurotransmitter release by a cochlear inner hair cell," J. Acoust. Soc. Am. **72**, 1435–1440.
Shannon, R. V., Zeng, F.-G., Wygonski, J., Kamath, V., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Siebert, W. M., and Gambardella, G. (**1968**). "Phenomenological model for a form of adaptation in primary auditory-nerve fibers," RLE QPR Communications Biophysics **88**, 330–334.
Smirnakis, S. M., Berry, M. J., Warland, D. K., Bialek, W., and Meister, M. (**1997**). "Adaptation of retinal processing to image contrast and spatial scale," Nature (London) **386**, 69–73.
Smith, R. L. (**1977**). "Short-term adaptation in single auditory-nerve fibers: Some post-stimulatory effects," J. Neurophysiol. **40**, 1098–1112.
Smith, R. L. (**1988**). "Encoding of sound intensity by auditory neurons," in *Auditory Function: Neurobiological Bases of Hearing*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 243–274.
Smith, R. L., and Brachman, M. L. (**1982**). "Adaptation in auditory-nerve fibers: A revised model," Biol. Cybern. **44**, 107–120.
Smith, R. L., Brachman, M. L., and Frisina, R. D. (**1985**). "Sensitivity of auditory-nerve fibers to changes in intensity: A dichotomy between decrements and increments," J. Acoust. Soc. Am. **78**, 1310–1316.
Smith, R. L., and Zwislocki, J. J. (**1975**). "Short-term adaptation and incremental responses of single auditory-nerve fibers," Biol. Cybern. **17**, 169–182.
Sumner, C. J., Lopez-Poveda, E. A., O'Mard, L. P., and Meddis, R. (**2002**). "A revised model of the inner-hair cell and auditory-nerve complex," J. Acoust. Soc. Am. **111**, 2178–2188.
Sumner, C. J., Lopez-Poveda, E. A., O'Mard, L. P., and Meddis, R. (**2003**). "Adaptation in a revised inner-hair cell model," J. Acoust. Soc. Am. **113**, 893–901.
Teich, M. C. (**1989**). "Fractal character of the auditory neural spike train," IEEE Trans. Biomed. Eng. **36**, 150–160.
Teich, M. C., and Lowen, S. B. (**1994**). "Fractal patterns in auditory nerve-spike trains," IEEE Eng. Med. Biol. Mag. **13**, 197–202.
Thorson, J., and Biederman-Thorson, M. (**1974**). "Distributed relaxation processes in sensory adaptation," Science **183**, 161–172.
Toib, A., Lyakhov, V., and Marom, S. (**1998**). "Interaction between duration of activity and time course of recovery from slow inactivation in mammalian brain Na^+ channels," J. Neurosci. **18**, 1893–1903.
Ulanovsky, N., Las, L., Farkas, D., and Nelken, I. (**2004**). "Multiple timescales of adaptation in auditory cortex neurons," J. Neurosci. **24**, 10440–10453.
Watkins, P. V., and Barbour, D. L. (**2008**). "Specialized neuronal adaptation for preserving input sensitivity," Norelco Rep. **11**, 1259–1261.
Westerman, L. A. (**1985**). "Adaptation and recovery of auditory nerve responses," Special Report No. ISR-S-24, Syracuse University, Syracuse, NY.
Westerman, L. A., and Smith, R. L. (**1987**). "Conservation of adapting components in auditory-nerve responses," J. Acoust. Soc. Am. **81**, 680–691.
Westerman, L. A., and Smith, R. L. (**1988**). "A diffusion model of the transient response of the cochlear inner hair cell synapse," J. Acoust. Soc. Am. **83**, 2266–2276.
Wixted, J. T., and Ebbesen, E. (**1997**). "Genuine power curves in forgetting," Mem. Cognit. **25**, 731–739.
Xu, Z., Payne, J. R., and Nelson, M. E. (**1996**). "Logarithmic time course of sensory adaptation in electrosensory afferent nerve fibers in a weakly electric fish," J. Neurophysiol. **76**, 2020–2032.
Yates, G. K. (**1987**). "Dynamic effects in the input/output relationship of auditory nerve," Hear. Res. **27**, 221–230.
Yates, G. K., Robertson, D., and Johnstone, B. M. (**1985**). "Very rapid adaptation in the guinea pig auditory nerve," Hear. Res. **17**, 1–12.
Young, E. D., and Sachs, M. B. (**1973**). "Recovery from sound exposure in auditory-nerve fibers," J. Acoust. Soc. Am. **54**, 1535–1543.
Zeddies, D. G., and Siegel, J. H. (**2004**). "A biophysical model of an inner hair cell," J. Acoust. Soc. Am. **116**, 426–441.
Zhang, X., and Carney, L. H. (**2005**). "Analysis of models for the synapse between the inner hair cell and the auditory nerve," J. Acoust. Soc. Am. **118**, 1540–1553.
Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (**2001**). "A phenomenological model for the responses of auditory-nerve fibers. I. Nonlinear tuning with compression and suppression," J. Acoust. Soc. Am. **109**, 648–670.
Zhang, F., Miller, C. A., Robinson, B. K., Abbas, P. J., and Hu, N. (**2007**). "Changes across time in spike rate and spike amplitude of auditory nerve fibers simulated by electric pulse trains," J. Assoc. Res. Otolaryngol. **8**, 356–372.
Zilany, M. S. A., and Bruce, I. C. (**2006**). "Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery," J. Acoust. Soc. Am. **120**, 1446–1466.
Zilany, M. S. A., and Bruce, I. C. (**2007**). "Representation of the vowel/in normal and impaired auditory nerve fibers: Model predictions of responses in cats," J. Acoust. Soc. Am. **122**, 402–417.
Zwicker, E. (**1984**). "Dependence of post-masking on masker duration and its relation to temporal effects in loudness," J. Acoust. Soc. Am. **75**, 219–223.

Contralateral acoustic stimulation alters the magnitude and phase of distortion product otoacoustic emissions

Ryan Deeter and Rebekah Abel
Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, Illinois 60208

Lauren Calandruccio
Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, Illinois 60208; Hugh Knowles Center for Hearing Research, Northwestern University, Evanston, Illinois 60208; and Department of Linguistics, Northwestern University, Evanston, Illinois 60208

Sumitrajit Dhar[a)]
Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, Illinois 60208 and Hugh Knowles Center for Hearing Research, Northwestern University, Evanston, Illinois 60208

(Received 17 June 2009; revised 6 August 2009; accepted 15 August 2009)

Activation of medial olivocochlear efferents through contralateral acoustic stimulation (CAS) has been shown to modulate distortion product otoacoustic emission (DPOAE) level in various ways (enhancement, reduction, or no change). The goal of this study was to investigate the effect of a range of CAS levels on DPOAE fine structure. The $2f_1$-f_2 DPOAE was recorded ($f_2/f_1=1.22$, $L_1=55$ dB, and $L_2=40$ dB) from eight normal-hearing subjects, using both a frequency-sweep paradigm and a fixed frequency paradigm. Contamination due to the middle ear muscle reflex was avoided by monitoring the magnitude and phase of a probe in the test ear and by monitoring DPOAE stimulus levels throughout testing. Results show modulations in both level and frequency of DPOAE fine structure patterns. Frequency shifts observed at DPOAE level minima could explain reports of enhancement in DPOAE level due to efferent activation. CAS affected the magnitude and phase of the DPOAE component from the characteristic frequency region to a greater extent than the component from the overlap region between the stimulus tones. This differential effect explains the occasional enhancement observed in DPOAE level as well as the frequency shift in fine structure patterns. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3224716]

PACS number(s): 43.64.Jb, 43.64.Bt, 43.64.Kc [BLM] Pages: 2413–2424

I. INTRODUCTION

The peripheral auditory system is primarily a receiver of acoustic signals and the gateway to higher processing centers. However, activity at this first way station is not a result of simple passive characteristics of the structures involved; an active "cochlear amplifier" modifies acoustic signals in specific and complex ways (Dallos, 2008). Otoacoustic emissions (OAEs), low-level signals produced in the cochlea, are an essential byproduct of the cochlear amplifier (Kemp, 1978).

Distortion product (DP) OAEs are evoked by simultaneous stimulation with two pure tones (f_1 and f_2, $f_1<f_2$), and are measured at frequencies that are arithmetic combinations of the stimulus frequencies. Although DPOAEs at several combination frequencies can be recorded from the normally functioning human ear, the DPOAE at $2f_1$-f_2 is the most extensively studied and used for clinical purposes (Gorga *et al.*, 1997).

DPOAEs are commonly measured at a few frequencies per octave, resulting in a general picture of overall cochlear function in that region. When recorded with higher frequency resolution, however, a pseudo-periodic pattern of peaks (maxima) and dips (minima) in DPOAE level as a function of frequency, known as fine structure, is routinely observed (e.g., Heitmann *et al.*, 1998; Talmadge *et al.*, 1999; Dhar *et al.*, 2002; Dhar and Shaffer, 2004). DPOAE fine structure is a consequence of constructive and destructive interferences between two components: one from the region where the activity patterns of the stimulus tones overlap on the basilar membrane, and the other from the characteristic frequency (CF) region of the DPOAE (Talmadge *et al.*, 1999). This detailed response allows further investigation of the cochlea within one region.

In this study, we are interested in the modulation of DPOAEs, and specifically DPOAE fine structure, by the medial olivocochlear (MOC) efferent system stimulated acoustically via the contralateral ear. The MOC fibers are thick, myelinated fibers that originate in the medial part of the superior olivary complex. The fibers project through the vestibular nerve and innervate the outer hair cells (OHCs) directly. Mediated by the neurotransmitter acetylcholine, MOC

[a)]Author to whom correspondence should be addressed. Electronic mail: s-dhar@northwestern.edu

neurons change the conductance and the stiffness of the OHC, leading to a general attenuation of cochlear amplifier gain (see Guinan, 2006 for a recent review). The effects of efferent-induced attenuation of the cochlear amplifier is evident in reduced basilar membrane vibrations (Murugasu and Russell, 1996) as well as altered auditory nerve fiber responses (Kawase *et al.*, 1993) and tuning (Guinan and Gifford, 1988).

With more invasive measurement techniques being impractical, OAEs have become an important tool for studying the effects of the MOC efferents in live humans. Mountain (1980) was among the first to demonstrate a reduction in OAE level following efferent stimulation. These results were later corroborated and extended by Siegel and Kim (1982), who also reported reduction, enhancement, or no change in OAE levels upon efferent activation. In many ways, we are still unraveling the underlying physiology linked to these diverse changes in OAE level due to efferent stimulation.

The early promise of gleaning clinically useful information from the modulation of OAEs by the efferent system has led to a search for the most robust and stable measure. In the case of DPOAEs in humans, the vast majority of reports have demonstrated a small reduction in DPOAE level (1–3 dB) due to the activation of the MOC efferent system (Bassim *et al.*, 2003; Abdala *et al.*, 2009). Enhancement of DPOAE level upon stimulation of the efferent system, much akin to the results of Siegel and Kim (1982), has been reported as well (Maison and Liberman, 2000; Müller *et al.*, 2005). While reduction in DPOAE level is easily explained as a consequence of attenuation of cochlear gain by the MOC efferents, a differential alteration of the two DPOAE components has been proposed as the cause behind the observed enhancement at certain frequencies. The observation of enhancement almost exclusively at minima in DPOAE level supports this theory (Zhang *et al.*, 2007). Minima are a result of cancellation due to phase opposition between the two DPOAE components (from the overlap and CF regions). Selective reduction in the CF component at such a frequency would "release" the overlap component from cancellation thereby causing an enhancement of the DPOAE level in the ear canal.

The bipolar effect of efferent stimulation on DPOAE level can be observed in the time domain as well. When the MOC efferents are stimulated while monitoring DPOAE level at a specific frequency, a reduction in DPOAE level is observed upon activation of the MOC efferents when the monitored frequency is a known level maximum. In contrast, an enhancement is observed when the monitored frequency is a known level minimum (Sun, 2008b).

Several measures of efferent-induced alteration of OAEs, the so-called MOC reflex, have been proposed in recent literature. Maison and Liberman (2000) measured a change in DPOAE level at a fixed frequency over time (~300 ms) following onset of the stimulus. Both reduction and enhancement were observed in response to varying stimulus level combinations. The authors dealt with this by designating the absolute value of the biggest change in an animal (guinea pig) as the MOC reflex strength. Absolute reflex values in this group of animals were distributed over a range of approximately 20 dB. Following a similar paradigm in a group of human subjects, using contralateral acoustic stimulation (CAS) of the efferent system (Müller *et al.*, 2005) recommended using the difference between the largest enhancement and the largest reduction as the MOC reflex metric. Since the largest differences between enhancement and reduction have been typically observed at frequencies of DPOAE level minima, recommendations for measuring the MOC reflex at frequencies of fine structure minima have also been made in the literature (Müller *et al.*, 2005; Wagner *et al.*, 2007).

A counterpoint to the recommendation of measuring the MOC reflex only at DPOAE level minima has been set forward in a group of very recent publications (Zhang *et al.*, 2007; Purcell *et al.*, 2008; Abdala *et al.*, 2009). While results from this set of experiments align with the previously reported robustness of the MOC reflex at DPOAE level minima, they also show extreme variability at such frequencies. These publications recommend measuring the MOC reflex strength at DPOAE level maxima if magnitude can be sacrificed for the sake of stability.

Using stimuli swept continuously in frequency yields a more comprehensive picture of the effects of MOC activation on DPOAE level (Purcell *et al.*, 2008; Abdala *et al.*, 2009). Purcell *et al.* (2008) reported the effect of the MOC efferents (stimulated acoustically via the contralateral ear) on DPOAE recordings with a fixed f_2 frequency while varying f_1 around f_2/f_1 ratios of either 1.10 or 1.22. Since this recording technique allows the comparison of a continuous DPOAE level versus frequency function with and without CAS, several metrics of the MOC reflex can be extracted, such as the maximum change at any given frequency, average change at all frequencies, or change in the area under the fine structure curve. These authors did not advocate for any particular metric but simply demonstrated the feasibility of using stimuli swept in frequency to evaluate multiple measures of the effect of MOC activation on DPOAEs. These data, dense in their frequency distribution, expose a consistent shift in DPOAE level toward higher frequencies upon activation of the MOC efferents.

A similar shift in DPOAE fine structure toward higher frequencies is also reported by other groups (Purcell *et al.*, 2008; Sun, 2008b; Abdala *et al.*, 2009). Abdala *et al.* (2009), in particular, focused on the effects of MOC activation on DPOAE fine structure and demonstrated that the level of the DPOAE CF component is affected more than that of the overlap component. Departing from the previous recommendation of measuring MOC effects at DPOAE level minima (Müller *et al.*, 2005; Wagner *et al.*, 2007), Abdala *et al.* (2009) explicitly recommended using the smaller but more stable reduction in DPOAE level observed at frequencies of DPOAE level maxima.

Acoustic stimuli used to activate the MOC efferents can also activate the middle ear muscle reflex (MEMR). The possibility of the MEMR co-occurring with the MOC reflex casts a constant shadow of doubt in the interpretation of data such as those reported here. A simple approach would be to stimulate the MOC efferents at levels lower than the MEMR threshold. However, the detected MEMR threshold is criti-

cally dependent on the limits of the measurement system and technique. Laboratory techniques (Guinan *et al.*, 2003; Goodman and Keefe, 2006) often yield thresholds significantly lower than those detected using commercially available impedance audiometers.

The relative contribution of the MEMR and the MOC reflex to changes in OAEs (or other measures of cochlear output, such as the compound action potential) appears to be species dependent. For example, the effect appears to be dominated by the MEMR in the rat, as sectioning the middle ear muscle eradicates the response almost entirely (Relkin *et al.*, 2005). In contrast, the effect appears to be dominated by the MOC efferents in the cat (Liberman *et al.*, 1996; Puria *et al.*, 1996) and humans (Giraud *et al.*, 1995). Nonetheless, the MEMR is an issue to contend with and we made a concerted effort to isolate the MOC reflex from the MEMR in the data reported here.

The quest to understand the effects of MOC efferents on OAEs in general, DPOAEs and DPOAE fine structure in particular, and to find a reliable and robust measure applicable for clinical use, is ongoing. Here we extend previous work by presenting results from eight normal-hearing healthy young adult humans, documenting the effects of multiple levels of CAS on DPOAEs. The subjects chosen for these experiments had unusually high MEMR but behavioral hearing thresholds within the normal range, and robust DPOAEs.

II. METHODS

Data from the right ears of eight normal-hearing female human subjects (mean age=20 years, st. dev.=1 year) are reported. These eight subjects, selected from a database of approximately 90 subjects, displayed pronounced DPOAE fine structure, modulation of DPOAEs due to contralateral stimulation (see top panel of Fig. 3 for example), and, most importantly, high MEMR thresholds. All tests were conducted in a sound-treated audiometric booth with the subject comfortably seated in a recliner. Subjects were paid for their participation and all measurements were conducted in accordance with the guidelines of the Institutional Review Board at Northwestern University.

A. General methods

Signal generation and recording for OAE measurements were done using custom software running on an Apple Macintosh computer. A MOTU 828 MkII input/output FireWire device was used for analog-to-digital (44 100 Hz, 24 bits) and digital-to-analog conversions. Generated signals were passed through a Behringer Pro XL headphone amplifier to MB Quart 13.01HX drivers. The drivers were coupled to the subjects' ear canal via an Etymotic Research ER10A probe assembly. The broadband noise (BBN) used for some experimental conditions was delivered through an independent channel of the MOTU and Behringer using an additional MB Quart driver. The signal from this driver was delivered to the ear canal using a plastic tube and foam tip assembly similar to one used in clinical audiometry. Signals from the test ear were recorded using the ER10A microphone and preamplifier combination, digitized using the MOTU and stored on disk for analyses.

Stimuli used for DPOAE measurements were calibrated using a coupler calibration procedure that allowed us to approximately compensate for the depth of insertion in the ear canal [Siegel (personal communication)]. The frequency responses of the transducers were measured in a long lossy tube (50 ft long, 0.375 in. outside diameter, copper plumbing tubing) using a slow chirp between 200 and 20 000 Hz. The absence of standing waves in this long tube with its diameter matched to that of the probe allows the recording of the combined frequency response of the sound source and the microphone. These frequency responses were also measured in an IEC 711 coupler (Bruel and Kjaer 4157) for various depths of insertion. The response was recorded using the microphone of the OAE probe as well as a Bruel and Kjaer 0.5-in. microphone (BK4134) attached at the distal end of the coupler. The recording obtained using the OAE probe microphone was normalized with that obtained in the long lossy tube at each insertion depth. This normalization resulted in the isolation of the frequency response of the cavity with a half-wave resonance that was related to the depth of insertion. The recording from the BK4134 was used to generate a correction filter for each insertion depth to yield a uniform stimulus level at the distal end of the coupler. During experiments, the frequency response for a specific insertion in each subject was measured in the ear canal using a slow chirp between 200 and 20 000 Hz. This response was normalized to that obtained in the long lossy tube and used to detect a half-wave resonance frequency. The compensation function for that particular half-wave frequency was then used to alter the stimuli before delivery to the ear canal.

B. Preliminary screening

The screening process consisted of otoscopy, pure-tone audiometry, tympanometry, measurement of contralateral acoustic reflex thresholds, and an OAE protocol described below. Audiometry was performed using standard clinical procedures (ASHA, 2005) bilaterally at 0.25, 0.5, 1, 2, 4, and 8 kHz using a Maico MI 26 clinical audiometer and TDH headphones. Tympanometry and contralateral reflex thresholds, using a 1000-Hz pure tone and a BBN as the activator in the non-test ear, were measured in the right ear (the test ear for all subjects) using an Interacoustic Audio Traveler AA220.

Spontaneous (S) OAEs and DPOAEs were measured during the screening procedure as well. A 3-min recording from the ear canal without any stimulation was analyzed using a 44,100-point fast Fourier transform (FFT) to detect SOAEs. The presence of SOAEs and their frequency location were noted and used in interpretation of results. Frequency regions with prominent SOAEs were not included in the DPOAE measurements.

DPOAEs were recorded by sweeping two stimulus tones between 500 and 6000 Hz [$L_1=55$ dB sound pressure level (SPL), $L_2=40$ dB SPL, and $f_2/f_1=1.22$] over 32 s (Long *et al.*, 2008). At least six sweeps were averaged before using a

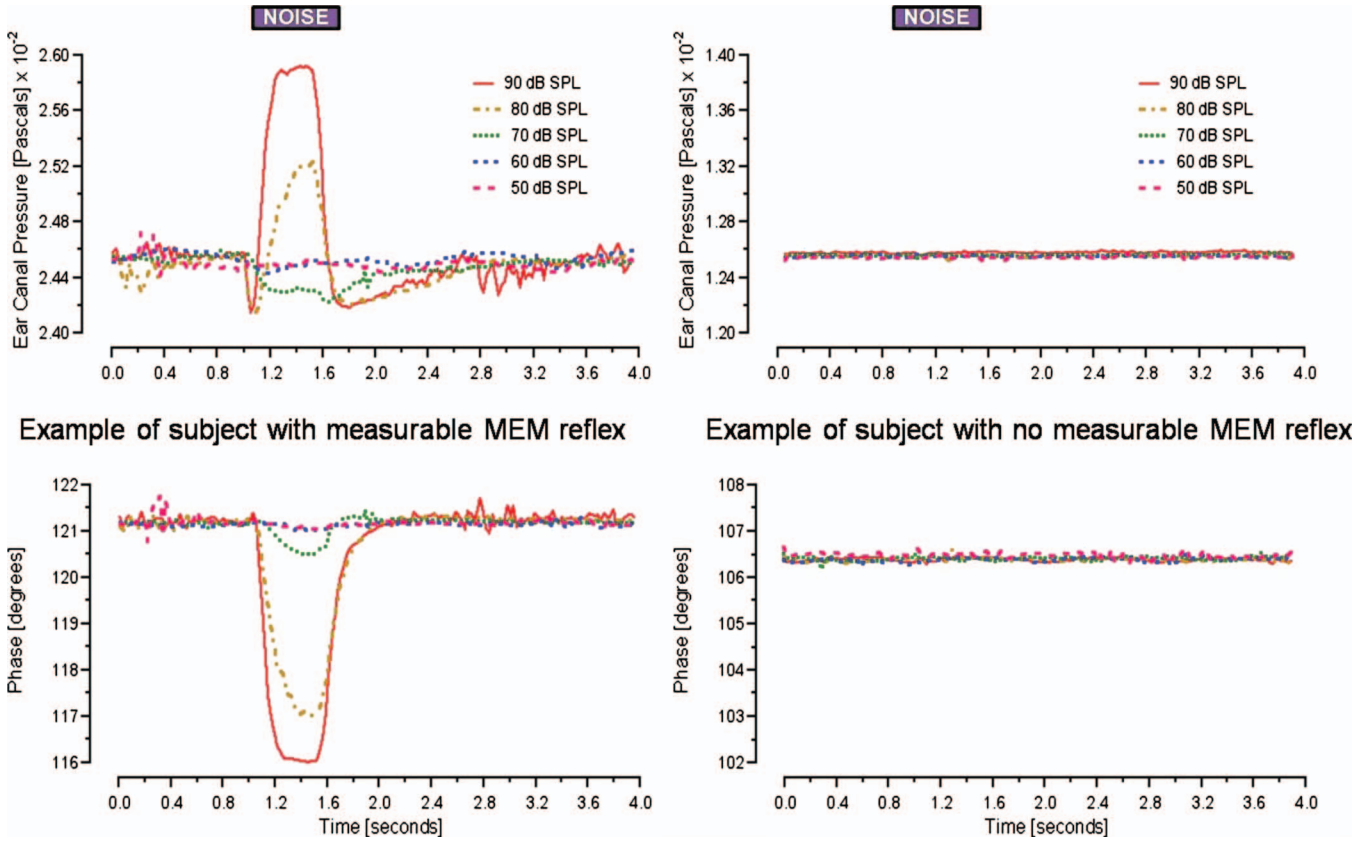


FIG. 1. Illustration of suspected MEMR recorded using laboratory technique. For all subjects a 602-Hz probe-tone was presented for 4 s in the test ear. A BBN activator was presented at five levels (50, 60, 70, 80, and 90 dB SPLs) in the contralateral ear between 1 and 1.5 s. Left panel: example of a subject showing measurable change in probe magnitude and phase starting at 70 dB SPL. Right: example of a subject with no measurable change in probe magnitude or phase at 90 dB SPL.

least-squares-fit (LSF) procedure (Long and Talmadge, 1997; Talmadge *et al.*, 1999; Dhar *et al.*, 2002, 2005) to estimate the level and phase of the DPOAE at the frequency $2f_1$-f_2. DPOAEs were also recorded over a narrower frequency range between 1000 and 2000 Hz (L_1=55 dB SPL, L_2=40 dB SPL, f_2/f_1=1.22, and stimuli swept over 8 s) with a BBN (0.1–10 kHz) presented in the contralateral ear at three levels (60, 70, and 80 dB SPL). The noise conditions were interleaved for level, with a 2-s silent interval between conditions, and were always preceded and followed by a recording without the contralateral noise. This entire sequence of five conditions was repeated eight times, allowing sufficient averaging to obtain an acceptable signal-to-noise ratio (SNR). The BBN was presented in the contralateral ear 1 s prior to the stimulus tones for all noise conditions to allow for the activation of the efferent response prior to initiation of DPOAE generation and recording. While this paradigm assures the activation of both fast and slow contralateral efferent effects, the stimulus tones themselves are likely to evoke ipsilateral efferent effects. Since our stimulus tones were swept in frequency always starting at the low frequencies, gradual activation of the slow ipsilateral efferent effects could affect the high frequencies more than the low frequencies. However, no frequency effects were found in our data leading us to collapse the results across frequencies in this report.

C. MEMR

To identify activation of the MEMR, the magnitude and phase of a 602-Hz probe-tone at 60 dB SPL were monitored in the test ear. A BBN (described above) was presented contralaterally in 10-dB steps between 50 and 90 dB SPLs. The noise was played for a period of 500 ms between 1 and 1.5 s, while the tone was monitored over a period of 4 s. Between 8 and 12 recordings were made at each noise level and averaged before extracting the magnitude and phase of the 602-Hz tone using a LSF analysis. See Fig. 1 (left panel) for an example of a subject where systematic changes in magnitude and phase of the 602-Hz tone induced by the contralateral noise were observed; we considered the increase in the magnitude of the 602-Hz tone, time-locked to the contralateral noise, as an indicator of activation of the MEMR. The subject in the left panel of Fig. 1 was not included in the current subject pool because her MEMR threshold was below our criterion (described below). In contrast, no significant and systematic changes in the magnitude and pressure of the 602-Hz tone were observed at any noise level for the subject in the right panel of Fig. 1. Results such as these were interpreted to indicate no significant middle ear involvement in response to contralateral noise stimulation up to 90 dB SPL.

All subjects included in these experiments had an MEMR threshold greater than 90 dB hearing loss (HL) for

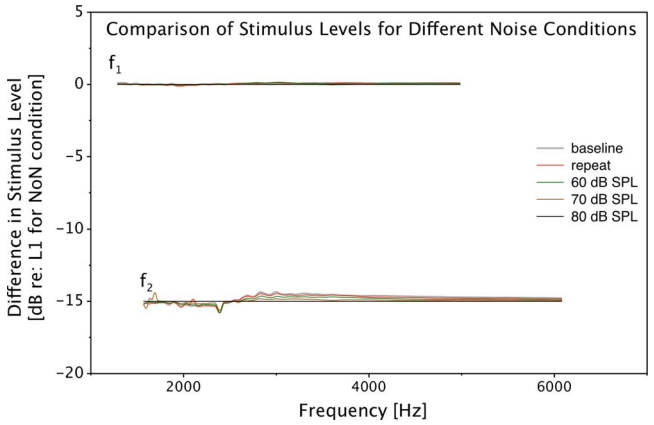


FIG. 2. (Color online) Comparison of stimulus levels for different test conditions plotted as the difference from the level of the low-frequency stimulus tone (f_1) for the NoN condition. The traces overlap signifying no systematic difference in the stimulus levels for the various noise conditions. Some separation between the f_2 levels can be observed. These changes were not found to be systematically related to the contralateral noise level.

both a 1-kHz tone and wide-band noise measured using the Interacoustics AA220. Three of our subjects showed observable changes in the magnitude and phase of the 602-Hz tone for a contralateral noise level of 90 dB SPL, while no changes in the magnitude and phase of the 602-Hz tone were observed at any noise level for the remaining five subjects. Other techniques for the same purpose have been reported in the literature (Guinan *et al.*, 2003; Goodman and Keefe, 2006) and such laboratory methods have typically been found to be more sensitive in detecting MEMRs than clinical methods. Our method of monitoring middle ear muscle activity was at least 20 dB more sensitive for these subjects than the thresholds obtained using the Interacoustics AA220.

A final control to ensure that the changes observed in DPOAEs due to contralateral stimulation were not dominated by the MEMR was implemented by comparing the levels of the stimulus tones across all test conditions. We hypothesized that if the contralateral noise induced a MEMR, the level of the stimulus tones in the ear canal would be higher as compared to the recording without contralateral stimulation. An example of such a comparison is presented in Fig. 2. As demonstrated by the virtual overlap of all traces in the figure, no significant or systematic differences in the levels of the stimulus tones were induced by the contralateral noise. Close examination of Fig. 2 reveals some variations in the level of f_2. However, these variations were not systematically related to the level of the contralateral noise. The level of f_2 is reduced below approximately 2100 Hz and increased beyond that. Further, the greatest deviation is observed for the no-noise (NoN) conditions. Given these idiosyncrasies, we attributed these deviations to errors in calibration and not to the MEMR. Results from all other subjects were indistinguishable from the example displayed in Fig. 2.

These results guided the establishment of an upper limit of 80 dB SPL for the contralateral noise presentation during the DPOAE recordings and allowed us to be reasonably certain that the changes observed in the DPOAE recordings were dominated by the efferent system.

D. Fixed frequency experiment

A second set of recordings was obtained at four fixed DPOAE frequencies in each subject. These data points were selected to represent DPOAE fine structure extrema (one maximum and one minimum from the same fine structure period in the NoN condition, two periods per subject). The stimuli were identical to those used in the DPOAE sweep measurements (i.e., L_1=55 dB SPL, L_2=40 dB SPL, f_2/f_1 =1.22, and DPOAE at $2f_1$-f_2) but with stimulus frequencies fixed to generate a DPOAE at the frequency of interest. Also, the same contralateral BBN conditions were used during these recordings (i.e., NoN; 60, 70, and 80 dB SPL; NoN for repeatability). Stimulus tones were played for 2 s, with the contralateral noise coming on at 0.5 s and turning off at 1 s with 5-ms on and off ramps. Thirty-two repetitions for each fixed frequency and each BBN condition were averaged.

E. Separation of DPOAE components

Components contributing to the DPOAE signal in the ear canal were separated using an inverse fast Fourier transform (FFT), time-windowing, and conversion to the frequency domain using an FFT (see Shaffer and Dhar, 2006 for methodological details). Briefly, subjecting the amplitude and phase of the DPOAE signal in the ear canal to an inverse FFT yields a pseudo-time domain representation of the DPOAE signal where the components from the overlap and DPOAE CF regions are distinguishable due to their differences in phase behavior as a function of frequency. Time domain filters were constructed for 400-Hz slices of data after visual inspection of the results of the inverse FFT. The two DPOAE components were then separated using these filters, and magnitude and phase of each component were individually reconstructed for the entire frequency range. Similar techniques have been used by several research groups in the literature (e.g., Stover *et al.*, 1996; Kalluri and Shera, 2001; Dhar *et al.*, 2005; Shaffer and Dhar, 2006).

F. Quantifying the effect of contralateral noise

The effect of varying levels of contralateral noise on DPOAE level and phase, as well as the level and phase of DPOAE components, was quantified using multiple measures. As has been the tradition, the effect of contralateral noise was computed as the difference in DPOAE level at a fixed frequency between the NoN condition and all noise conditions. This computation was done on three selected DPOAE level maxima and minima in each subject. These data points were selected such that at least a 3-dB SNR was maintained even for the highest level of contralateral noise. The DPOAE levels reported at an SNR of 3 dB could be influenced by external noise. However, choosing a higher SNR would not have allowed the analysis of DPOAE level minima. In parallel, the change in DPOAE level at these maxima and minima was also computed by tracking the maxima or minima as they shifted to different frequencies due to the contralateral noise. Finally, the shift in frequency of these maxima and minima was computed.

The change in level and phase of DPOAE components due to the contralateral noise was computed from the values

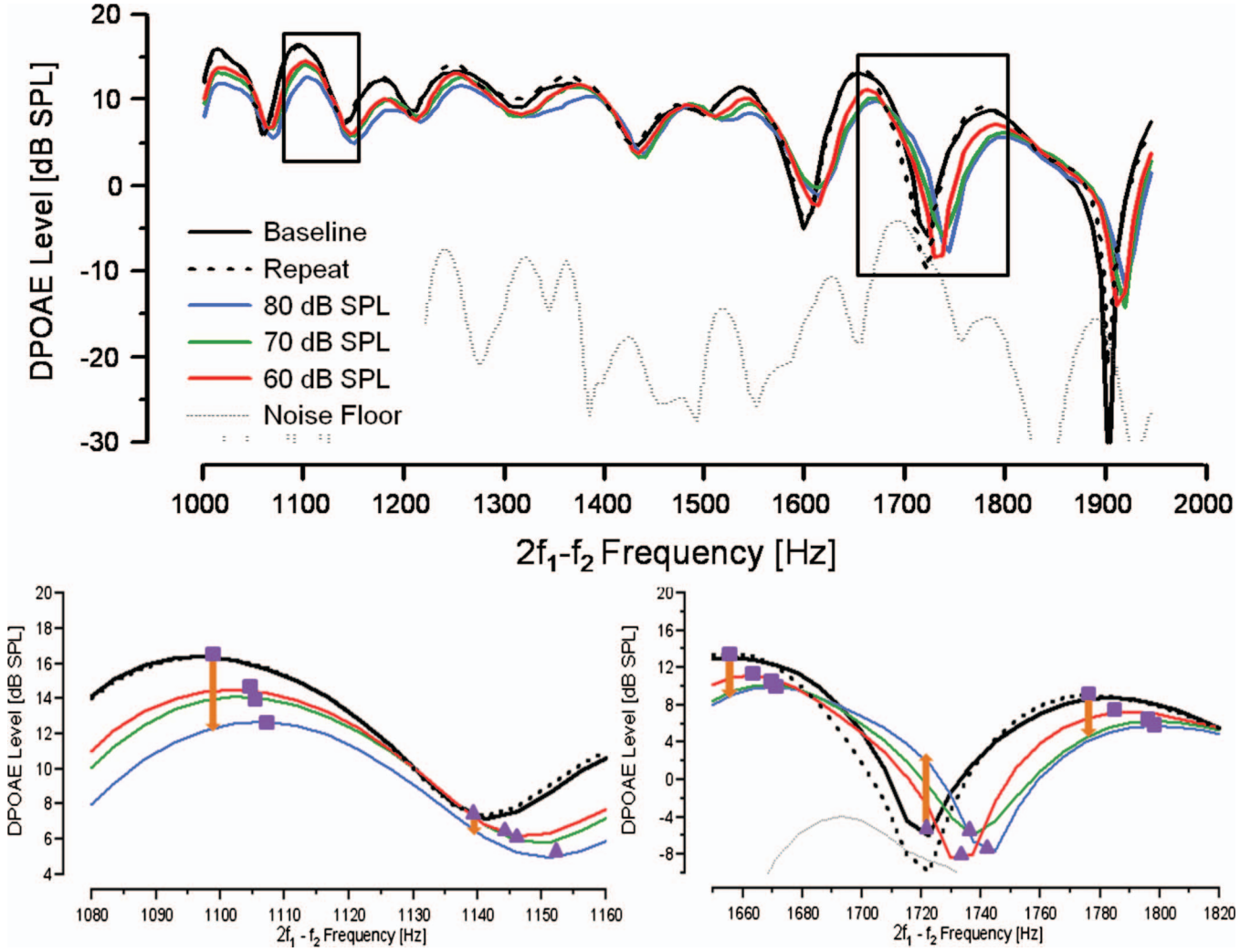


FIG. 3. Top panel: example of changes in DPOAE fine structure in response to several levels of CAS in an individual subject. Black boxes surround areas magnified in lower panels. Bottom panels: magnified view of fine structure extrema under different noise conditions. Purple squares and triangles represent maxima and minima, respectively. Orange arrows indicate the trend that traditional fixed frequency DPOAE analysis would have shown. Note that the minimum in the bottom left panel shows reduction (arrow pointing downwards) while the minimum in the bottom right panel shows enhancement (arrow pointing upwards).

for the NoN condition. These differences across frequency were then condensed to averages over third octave bands between 1 and 2 kHz. Statistical comparisons were made using SPSS version 17.0 (SAS, 2008).

III. RESULTS

A. General effects of contralateral noise

A general reduction in DPOAE level, an attenuation of fine structure depth, and a shift of fine structure patterns toward higher frequencies were observed for all three levels (60, 70, 80 dB SPLs) of contralateral BBN. These effects were more pronounced as the level of the contralateral noise increased. DPOAE level as a function of frequency from one subject (MOC038) for the initial and repeated NoN conditions (black traces), as well as for the three different noise conditions (color traces), is plotted in the top panel of Fig. 3 for illustrative purposes. Trends observed in Fig. 3 were consistent across subjects. The bottom two panels of Fig. 3 provide magnified views of two frequency regions enclosed in black boxes in the top panel. The fine structure extrema (maxima and minima) are tracked using purple squares and triangles, respectively. This tracking illustrates the shift of fine structure patterns toward higher frequencies.

The effects of contralateral BBN on DPOAE level were quantified in two ways. First, consistent with traditional analyses, the change in level at a fixed frequency was computed for each maximum or minimum. The orange arrows in the bottom panels of Fig. 3 schematically represent such a computation. Second, the change in DPOAE level for a given maximum or minimum was computed by tracking the level of that particular maximum or minimum, marked by the purple squares and triangles in the bottom panels of Fig. 3. Comparing the bottom panels of Fig. 3 clearly demonstrates the divergent results obtained when using the two different methods. The level of the minimum in the bottom right panel of Fig. 3 shows no systematic change with increasing contralateral BBN levels if the level at the exact minimum is tracked. However, an enhancement would be reported using the more traditional, fixed frequency method

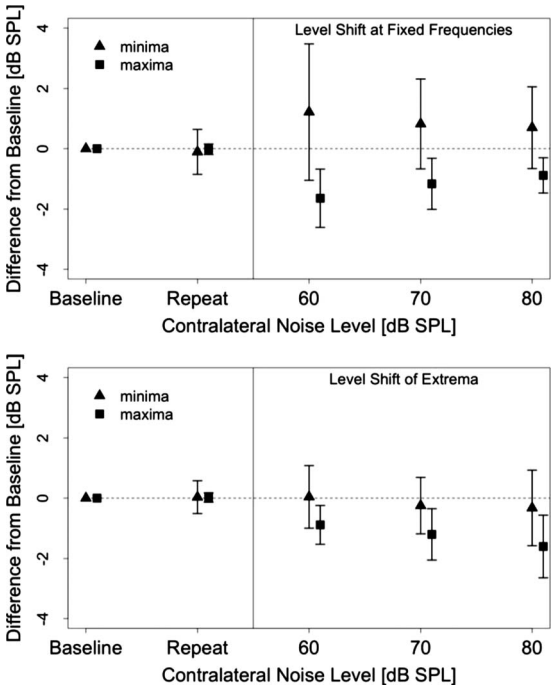


FIG. 4. Average level shifts at DPOAE fine structure extrema as a function of contralateral noise level at fixed frequencies (top panel) or for tracked extrema (bottom panel). Triangles represent fine structure minima and squares represent fine structure maxima. The DPOAE level recorded in the NoN condition established the baseline from which the shifts were calculated. A second NoN condition was included for test/retest reliability. Error bars represent one standard deviation. Three maxima and three minima from each of eight subjects (24 total points) were averaged.

of comparison (orange arrow). In contrast, both methods would yield a reduction in DPOAE level in the case of the minimum in the bottom left panel of Fig. 3. These divergent results in two frequency regions within the same subject are due, primarily, to the shift in fine structure frequency. For a similar shift in frequency, the sharper minimum around 1720 Hz shows an apparent enhancement at the minimum frequency as compared to the reduction observed at the minimum around 1095 Hz.

Three DPOAE level maxima and three minima were sampled from the eight subjects (total of 24 data points) where an SNR of at least 3 dB was maintained for all noise conditions. Maxima were chosen from the same fine structure period after an appropriate minimum was identified. Average changes at DPOAE level maxima (squares) and minima (triangles) computed using the traditional, fixed frequency method (top panel) and the tracking method (bottom panel) are presented in Fig. 4. The error bars represent one standard deviation. The averages for the baseline or NoN condition fall on the zero line by construction. Averages for the repeated NoN condition are also presented and show no differences from the baseline measures. DPOAE level at maxima shows a general reduction for the noise conditions using either method of calculation. The magnitude of reduction systematically increases with increasing noise level when the maximum frequency is tracked. The results are not as consistent for the fixed frequency method.

On average, DPOAE level minima show enhancement when the effect of contralateral noise is computed using the

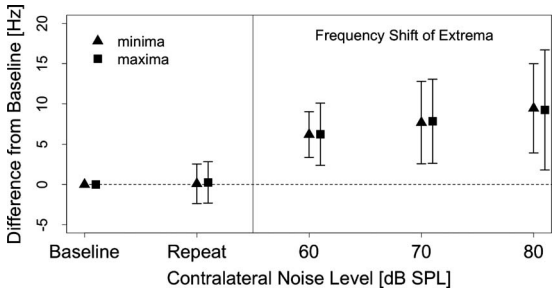


FIG. 5. Average frequency shifts at DPOAE fine structure extrema as a function of contralateral noise level. Triangles and squares represent minima and maxima, respectively. The frequency of a given extrema in the NoN condition established the baseline from which the frequency shifts were calculated. A second NoN condition was included for test/retest reliability. Error bars represent one standard deviation. Three maxima and three minima from each of eight subjects (24 total points) were averaged.

fixed frequency method (top panel of Fig. 4). In contrast, small reductions are observed in the average level of minima in the bottom panel of Fig. 4. Greater variability is observed in the minima for either method of computation, as compared to the maxima. However, the variability in minima is observably higher for the fixed frequency calculation.

Average shifts in frequency for minima and maxima are presented in Fig. 5. The format of the figure is similar to the panels in Fig. 4. On average, the frequencies of maxima and minima appear to be consistent across the two NoN conditions. Both minima and maxima show a consistent shift toward the higher frequencies with increasing contralateral noise levels. The variability for both maxima and minima appears to increase with increasing contralateral noise levels.

B. Effects on DPOAE components

The ear canal DPOAE was decomposed into the constituent components from the overlap and CF regions. Figure 6 displays DPOAE level as a function of frequency (top row) and phase (bottom row) of each DPOAE component from subjects MOC013 (left column) and MOC089 (right column). The range of the ordinate and the relatively small differences in DPOAE component phase for different noise conditions make it difficult to distinguish different traces in the bottom panels. The insets in the bottom panels highlight the phase behavior of each component over a 10-Hz frequency range. The bars next to the insets span a half cycle range. The phase of the overlap component is relatively invariant with frequency in both subjects. In contrast, the phase of the CF component accumulates approximately 6–8 cycles in the frequency span displayed.

Changes in the phase of the overlap component due to contralateral noise are difficult to detect in either subject. The phase of the CF component shows a visually detectable lead without a change in slope in the inset for subject MOC013 (bottom left) for the noise conditions. This was the pattern observed in seven of the eight subjects. The only exception to this pattern was seen in subject MOC089 (bottom right). In this case the slope of the phase of the CF component was steeper in the presence of contralateral noise between 1100 and 1300 Hz. While a consistent shift of fine structure toward higher frequencies is observed in subject MOC013 (left

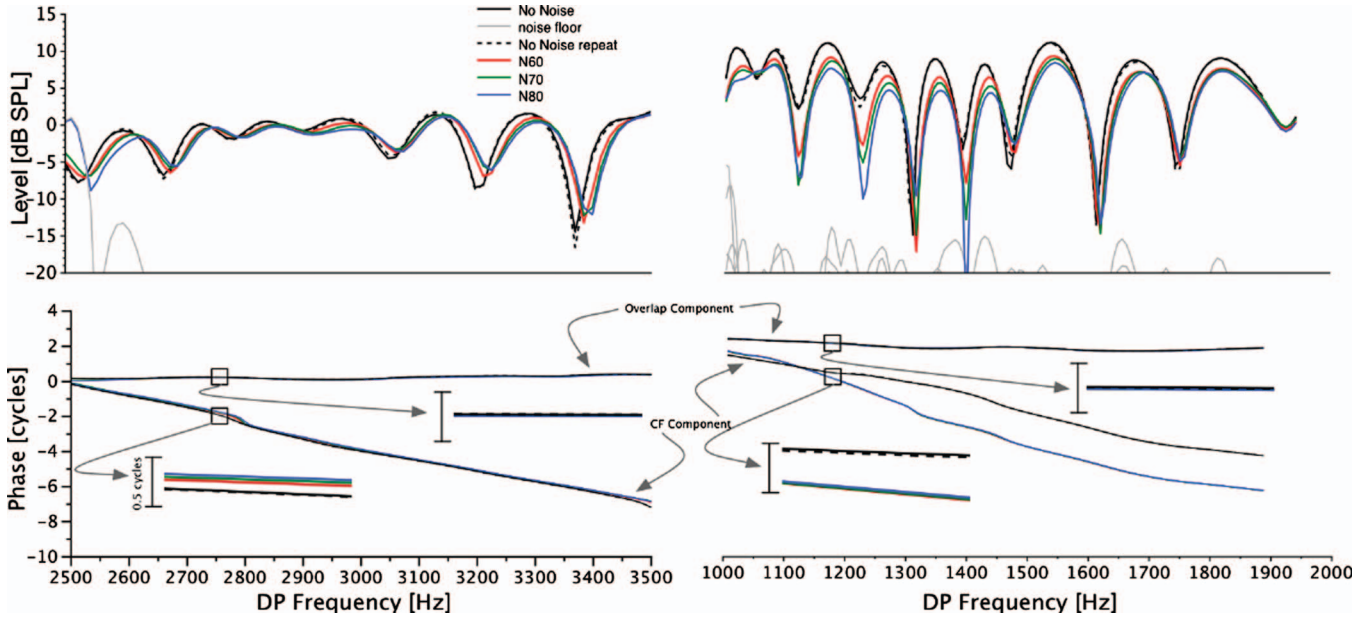


FIG. 6. Ear canal DPOAE level (top) and DPOAE component phases (bottom) as a function of frequency in two subjects. See text for method used to separate DPOAE components. The insets in the bottom panel offer a detailed view of the phases of each component over a 10-Hz range with different levels of contralateral noise as the parameter. The typical response of a larger change in the phase of the CF component is observed in the bottom left panel. The bottom right panel depicts an atypical response, in only one of eight subjects, where the slope of the phase of the CF component was altered in the presence of contralateral noise.

column), such a shift, especially between 1100 and 1300 Hz, is not observed in subject MOC089 (right column). We emphasize that the pattern seen in the bottom right panel was unique to subject MOC089; data from all other subjects showed a pattern consistent with that observed for subject MOC013 in the left column.

Changes in DPOAE component level (top) and phase (bottom) averaged across the entire frequency range and then averaged across subjects are displayed in Fig. 7 as a function of BBN level. The error bars represent one standard deviation from the mean. Greater reduction in the magnitude of the CF component and greater increase in the phase of the CF component were observed for all noise conditions. Paired *t*-tests indicated significant differences in both magnitude and phase between the two DPOAE components ($p<0.008$, compensated for multiple comparisons) for all but one paired contrast: The change in phase of the two components was not statistically significantly different for the contralateral noise level of 80 dB SPL [$t(373)=2.601, p=0.01$]. We draw the reader's attention to the greater reduction in the level and greater increase in the phase of the CF component. Much of our interpretation of the data will hinge on this.

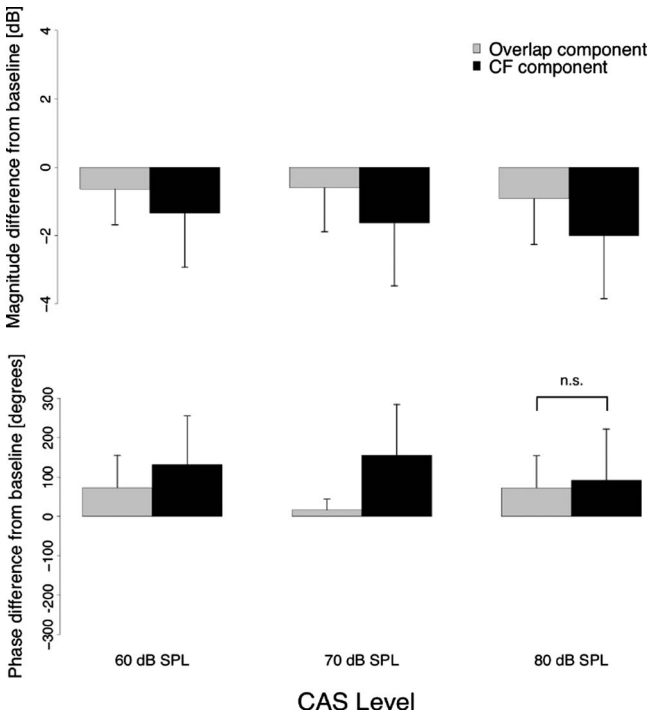


FIG. 7. Average change in level (top) and phase (bottom) in DPOAE components due to different levels of contralateral noise. The reported differences were calculated by first averaging the component level or (unwrapped) phase over the entire recorded frequency range and then subtracting the average values of the NoN condition from the average values for each noise condition. Error bars represent one standard deviation. Top panel: change in each component level for three different noise levels. Bottom panel: change in each component phase for three different noise levels. All contrasts between the two components were statistically significant ($p<0.05$) except the change in phase of the two components for the contralateral noise level of 80 dB SPL.

C. Results at fixed frequencies

To confirm the results observed in the data obtained with swept stimulus frequencies, DPOAE levels were monitored over time at fixed frequencies corresponding to DPOAE level maxima or minima while presenting different levels of contralateral noise. Results of this experiment from one subject at one minimum (top) and one maximum (bottom) are presented in Fig. 8. The time window in which the contralateral noise was presented is marked in the figure. In the case of this minimum, DPOAE level increased due to the con-

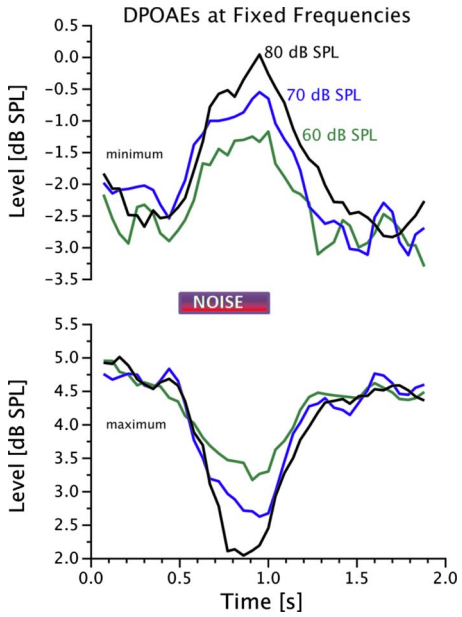


FIG. 8. (Color online) Change in DPOAE level at a fixed frequency over time with different levels of contralateral noise presented for 500 ms between 0.5 and 1 s. Recordings were made for a fine structure minimum (top panel) or maximum (bottom panel) as determined from the NoN condition.

tralateral noise, and this increase was more pronounced for greater levels of contralateral noise. In the case of this maximum, the level of the DPOAE showed a consistent reduction due to the contralateral noise, and the reduction was more pronounced with increasing noise levels. While all 24 DPOAE level maxima monitored in this experiment showed consistent reduction, the effects of contralateral noise were more varied for the minima.

Examples of the three types of effects observed at minima are displayed in Fig. 9. An example of contralateral noise-induced enhancement is displayed in the left panel. This effect was observed in 6 of the 24 minima. The most common observation, in 14 of the 24 minima, was of no consistent change in DPOAE level, as shown in the middle panel. Finally, reduction in DPOAE level upon presentation of contralateral noise as shown in the right panel was observed in 4 of the 24 minima. When either enhancement or reduction was observed, the effect grew with increasing noise level. The time course of change in DPOAE level displayed in Figs. 8 and 9 should not be interpreted to be physiologically relevant as they were influenced by the averaging employed to enhance SNR.

IV. DISCUSSION

The results presented here continue the examination of the modulation of OAEs by the olivocochlear efferents that was started almost 3 decades ago (Mountain, 1980; Siegel and Kim, 1982). Most recently, an in-depth examination of efferent alteration of DPOAE fine structure has become one of the foci of this area of work (Zhang *et al.*, 2007; Purcell *et al.*, 2008; Sun, 2008a, 2008b; Abdala *et al.*, 2009). We begin by highlighting the consistency of these results with previous work and then discussing the issues of the frequency shift observed in fine structure patterns and the most appropriate clinical measure of the efferent control of DPOAEs.

A. Reduction and enhancement of DPOAE level

The change in DPOAE level at fixed frequencies is the appropriate comparison with data from literature. Consistent with previous work (e.g., Lisowska *et al.*, 2002; Zhang *et al.*, 2007; Abdala *et al.*, 2009), an average reduction of 1.64 ± 0.96 dB at maxima and average enhancement of 1.21 ± 2.83 dB at minima were observed for the contralateral noise condition of 60 dB SPL. For the 70 dB SPL condition, the average level shifts were 1.16 ± 0.85 dB (reduction) at maxima and 0.82 ± 1.40 dB (enhancement) at minima, also consistent with previous work (Zhang *et al.*, 2007). The magnitude of the effects continued to decline at both maxima and minima (Fig. 4) for the 80 dB SPL condition. The consistency of these results with previous work survives the methodological differences between these reports. For example, Abdala *et al.* (2009) used stimulus levels of 65 and 55 dB SPLs and reported efferent-induced reduction in level at maxima averaged across frequency bands spanning 2 mm of the basilar membrane.

Both minima and maxima showed a more consistent change with increasing noise levels when the effects of the contralateral noise were quantified by tracking individual extrema (Fig. 4, bottom panel). Changes were smaller in magnitude at the minima using this method, but the average change was a reduction rather than an enhancement. The magnitude of the effect exhibited growth for both maxima and minima with increasing contralateral noise level. The effect was smaller for the minima as compared to the maxima, arguably due to the limitation imposed by the noise floor in the measurement ear.

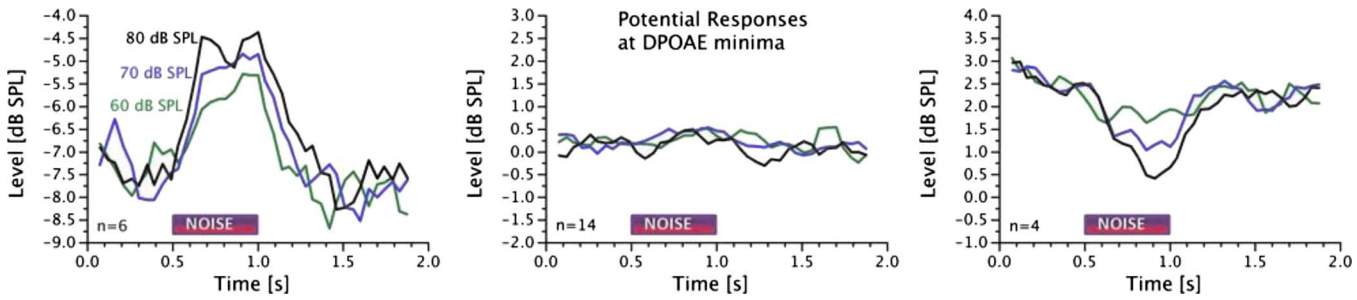


FIG. 9. (Color online) Examples of three types of responses observed at fine structure minima in the fixed frequency experiment. Upon the presentation of the contralateral noise, DPOAE level increased in 6 cases, did not show systematic change in 14 cases, and decreased in 4 cases of the 24 minima examined.

Also consistent with literature, the effect of the contralateral noise was variable at minima, as evident in the larger variance in Fig. 4 and the variety of effects seen in the time domain in Fig. 9. Such variability has been reported by others (Zhang *et al.*, 2007) with the greatest effect often observed at these minima, especially when the absolute value of the change (irrespective of the direction) is considered (Müller *et al.*, 2005; Wagner *et al.*, 2007).

An important question that many authors have addressed is that of the most appropriate *clinical* measure of the effects of the efferent system on DPOAEs. Some have argued for using the measure that yields the largest response, often seen at minima, especially when the range of change from the greatest reduction to the greatest enhancement is taken into account (Müller *et al.*, 2005; Wagner *et al.*, 2007). Others have supported the use of a more consistent (but smaller) effect by measuring the change only at maxima (e.g., Abdala *et al.*, 2009). Yet others have offered an array of quantification schemes available from swept-frequency measurements such as those employed here (Purcell *et al.*, 2008). In a similar vein, we have presented two alternate methods of quantifying the effects of efferent activation on DPOAEs, first by measuring the effect at a fixed frequency, and second by tracking individual maxima and/or minima as they shift in frequency with the introduction of contralateral noise. In either case, data at maxima appear to yield more consistent results.

Given the capacity of current clinical equipment, it is unrealistic to expect a viable protocol where the clinician first searches for a DPOAE level maximum in a given frequency range and then measures the effects of contralateral stimulation at that target frequency. A more panoramic measure using stimulus tones swept continuously in frequency as employed here and elsewhere (Purcell *et al.*, 2008; Abdala *et al.*, 2009) is not available to the clinician either. An alternate approach could be to sample each fine structure period by making measurements at a few adjacent target frequencies and then using the largest reduction as the nominal measure of efferent alteration of DPOAE. Spontaneous, transient evoked, and stimulus frequency OAEs show a consistent spacing pattern that can be characterized to the value $f/\Delta f = 16$, where f is the center frequency and Δf is the frequency distance between adjacent SOAEs or maxima/minima of transient or stimulus frequency OAEs (e.g., Zwicker, 1990; Shera, 2003). Although shown not to coincide exactly with other types of OAEs (Lutman and Deeks, 1999), approximate correspondence of DPOAE fine structure spacing to these other measures has been demonstrated (Reuter and Hammershoi, 2006; Dhar and Abdala, 2007). Based on this estimate of fine structure spacing, measurements made at f, $f \pm (f/4)$, and $f \pm (f/8)$, where f is the frequency of interest, should allow the sampling of a data point at or near the maximum of that particular fine structure period. Thus, measuring the effect of the MOC on five frequency points would give a realistic chance of sampling one frequency near or at a DPOAE level maximum. This data point could be identified as the one showing a small but consistent effect on multiple measures and then be used as the metric of MOC strength in that frequency range for that subject.

B. The second dimension: A shift in frequency

A key observation of this study is that contralateral noise causes a frequency shift in DPOAE fine structure. Such a shift in frequency has been observed (Purcell *et al.*, 2008; Sun, 2008b) and quantified (Abdala *et al.*, 2009) in recent reports. The magnitude and direction of frequency shifts observed here are consistent with previous reports. Both maxima and minima show consistent and parallel shifts in frequency (Fig. 5). Shifts in spontaneous emission frequency due to CAS have been reported before (Mott *et al.*, 1989; Harrison and Burns, 1993) and are arguably mediated by similar mechanisms that cause shifts in DPOAE fine structure.

Although the shift in DPOAE fine structure has been demonstrated before, the cause behind this shift has not been explored in detail. Abdala *et al.* (2009) (p. 1593) commented the following: "Future studies are warranted to further elucidate the mechanism and biological significance of this frequency shift." We are able to demonstrate at least the first layer of cochlear macromechanics responsible for this frequency shift.

It has long been speculated (Siegel and Kim, 1982; Maison and Liberman, 2000; Kim *et al.*, 2001; Zhang *et al.*, 2007; Purcell *et al.*, 2008; Sun, 2008a, 2008b) and more recently documented (Abdala *et al.*, 2009) that the activation of the efferents has a differential effect on the two DPOAE components, with the component from the CF region being affected more than the component from the overlap region. However, this greater effect has been demonstrated for component levels only. That is, the level of the CF component is reduced more than the level of the overlap component upon efferent stimulation (Abdala *et al.*, 2009). Such a model can fully explain the enhancement of the DPOAE level observed in the ear canal at DPOAE level minima. The two DPOAE components are at phase opposition at minima and selective or greater reduction in the CF component would in effect release the overlap component from cancellation and cause an enhancement in the overall signal in the ear canal. A frequency shift in fine structure patterns is not as easily explained by this differential change in DPOAE component level. However, these and previous results (Zhang *et al.*, 2007; Purcell *et al.*, 2008; Sun, 2008a, 2008b; Abdala *et al.*, 2009) definitively indicate that frequency shifts in fine structure patterns are at least as responsible as the release of the overlap component from cancellation for the enhancements observed at DPOAE minima due to MOC activation.

However, the frequency shift can be explained if the phases of the two DPOAE components are affected differently by efferent activation. One possibility is that the slope of the phase of the CF component is changed by the contralateral noise, while that of the overlap component is either unchanged or changed to a relatively lesser degree. In order to result in a shift toward the higher frequencies, the slope of the CF component would have to become shallower upon presentation of the contralateral noise relative to the NoN condition. This would result in a shift of DPOAE frequency toward higher frequencies as well as a widening of fine structure spacing. Fine structure spacing has been reported to re-

main unchanged upon efferent activation (Abdala *et al.*, 2009), a result supported observationally in our data even though we did not quantify fine structure spacing.

A shift toward the higher frequencies can also be caused if the overall phase of the CF component shows a lead due to the contralateral noise. For a frequency shift to occur, this lead has to be greater than any lead seen in the phase of the overlap component. This appears to be the case as displayed in the bottom panel of Fig. 7. While both the overlap and CF component phases show a lead at all levels of contralateral noise, the change is statistically significantly larger for the CF component (except for the noise condition at 80 dB SPL). This shift in phase without a change in the slope of the phase results in a shift of DPOAE fine structure toward higher frequencies without a change in spacing. A change in the slope of the CF component phase was observed in one subject (MOC089, right panel of Fig. 6). The slope of the CF component phase became steeper with the introduction of the noise but frequency shifts in the fine structure pattern were not evident, especially between 1100 and 1300 Hz where the change in slope was most evident.

C. MOC effect and MEM contamination

All reports of modulation of OAEs by acoustic stimulation of the efferents must carefully treat the issue of accidental activation of the MEMR. We attempted to control for this by choosing a group of subjects with high MEMR thresholds and then making additional measurements to ensure that the MEMR was not activated for the levels of contralateral noise used (see Figs. 1 and 2). Based on these precautions, we are fairly confident that the effects demonstrated here are dominated by the MOC efferents. Consistent with this assertion, we did not observe any abrupt discontinuities in the trajectories of change in either DPOAE level or frequency with increasing stimulus level. Such a change in trajectory has been associated with the activation of the MEMR (Guinan *et al.*, 2003).

V. CONCLUSIONS

Contralateral stimulation with BBN causes changes in both DPOAE level and frequency of fine structure patterns. The most consistent reduction is observed by tracking the change in DPOAE level maxima, taking into account the shifts in the frequency of these patterns with the introduction of contralateral noise. The oft-reported enhancement at minima is caused not only by a greater reduction in the DPOAE component from the CF region but also by a shift in DPOAE fine structure patterns toward higher frequencies. Both the level and the phase of the CF component are affected to a greater extent by the contralateral noise. The effective input for the generation of the CF component is the portion of the overlap component that travels inward rather than outward toward the ear canal. This input is smaller in magnitude than the stimulus tones that serve as the input for the generation of the overlap component. Whether this greater effect on the CF component is because of a lower level input or reflective of mechanistic differences in the generation of the two DPOAE components remains an open question. Since both the level and phase of DPOAE components are affected by the contralateral noise, tracking true change in DPOAE level is best achieved by tracking the magnitude of a level maximum or minimum as it shifts in frequency. This ensures that the observed change in DPOAE magnitude is due mostly to changing magnitudes of DPOAE components as changes in the relative phase of the two components are accounted for.

ACKNOWLEDGMENTS

We would like to thank Dashiell Oatman-Stanford and Deepika Sriram for assistance with data collection. Helpful discussions with Wei Zhao, Renee Banakis, Hwa Jung Son, Mead Killion, Carolina Abdala, and Jonathan Siegel shaped our thinking about the issues presented here. This research was funded by Grant Nos. R01DC008420 and R01DC003552 (subcontract to Northwestern from the House Ear Institute) from the NIDCD, and the Hugh Knowles Center for Hearing Research at Northwestern University. Parts of this work were presented at the 2009 Annual Convention of the American Auditory Society.

Abdala, C., Mishra, S. K., and Williams, T. L. (**2009**). "Considering distortion product otoacoustic emission fine structure in measurements of the medial olivocochlear reflex," J. Acoust. Soc. Am. **125**, 1584–1594.

American Speech-Language-Hearing Association (ASHA) (**2005**). *Guidelines for Manual Pure-Tone Threshold Audiometry*, ASHA, Rockville, MD.

Bassim, M. K., Miller, R. L., Buss, E., and Smith, D. W. (**2003**). "Rapid adaptation of the 2f1-f2 DPOAE in humans: Binaural and contralateral stimulation effects," Hear. Res. **182**, 140–152.

Dallos, P. (**2008**). "Cochlear amplification, outer hair cells and prestin," Curr. Opin. Neurobiol. **18**, 370–376.

Dhar, S., and Abdala, C. (**2007**). "A comparative study of distortion-product-otoacoustic-emission fine structure in human newborns and adults with normal hearing," J. Acoust. Soc. Am. **122**, 2191–2202.

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

Dhar, S., and Shaffer, L. A. (**2004**). "Effects of a suppressor tone on distortion product otoacoustic emissions fine structure: Why a universal suppressor level is not a practical solution to obtaining single-generator DP-grams," Ear Hear. **25**, 573–585.

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

Giraud, A. L., Collet, L., Chery-Croze, S., Magnan, J., and Chays, A. (**1995**). "Evidence of a medial olivocochlear involvement in contralateral suppression of otoacoustic emissions in humans," Brain Res. **705**, 15–23.

Goodman, S. S., and Keefe, D. H. (**2006**). "Simultaneous measurement of noise-activated middle-ear muscle reflex and stimulus frequency otoacoustic emissions," J. Assoc. Res. Otolaryngol. **7**, 125–139.

Gorga, M. P., Neely, S. T., Ohlrich, B., Hoover, B., Redner, J., and Peters, J. (**1997**). "From laboratory to clinic: A large scale study of distortion product otoacoustic emissions in ears with normal hearing and ears with hearing loss," Ear Hear. **18**, 440–455.

Guinan, J. J., Jr. (**2006**). "Olivocochlear efferents: Anatomy, physiology, function, and the measurement of efferent effects in humans," Ear Hear. **27**, 589–607.

Guinan, J. J., Jr., Backus, B. C., Lilaonitkul, W., and Aharonson, V. (**2003**). "Medial olivocochlear efferent reflex in humans: Otoacoustic emission (OAE) measurement issues and the advantages of stimulus frequency OAEs," J. Assoc. Res. Otolaryngol. **4**, 521–540.

Guinan, J. J., Jr., and Gifford, M. L. (**1988**). "Effects of electrical stimulation of efferent olivocochlear neurons on cat auditory-nerve fibers. III. Tuning curves and thresholds at CF," Hear. Res. **37**, 29–45.

Harrison, W. A., and Burns, E. M. (**1993**). "Effects of contralateral acoustic stimulation on spontaneous otoacoustic emissions," J. Acoust. Soc. Am.

94, 2649–2658.
Heitmann, J., Waldmann, B., Schnitzler, H. U., Plinkert, P. K., and Zenner, H. P. (**1998**). "Suppression of distortion product otoacoustic emissions (DPOAE) near f1-f2 removes DP-gram fine structure—Evidence for a secondary generator," J. Acoust. Soc. Am. **103**, 1527–1531.
Kalluri, R., and Shera, C. A. (**2001**). "Distortion-product source unmixing: A test of the two-mechanism model for DPOAE generation," J. Acoust. Soc. Am. **109**, 622–637.
Kawase, T., Delgutte, B., and Liberman, M. C. (**1993**). "Antimasking effects of the olivocochlear reflex. II. Enhancement of auditory-nerve response to masked tones," J. Neurophysiol. **70**, 2533–2549.
Kemp, D. T. (**1978**). "Stimulated acoustic emissions from within the human auditory system," J. Acoust. Soc. Am. **64**, 1386–1391.
Kim, D. O., Dorn, P. A., Neely, S. T., and Gorga, M. P. (**2001**). "Adaptation of distortion product otoacoustic emission in humans," J. Assoc. Res. Otolaryngol. **2**, 31–40.
Liberman, M. C., Puria, S., and Guinan, J. J., Jr. (**1996**). "The ipsilaterally evoked olivocochlear reflex causes rapid adaptation of the 2f1-f2 distortion product otoacoustic emission," J. Acoust. Soc. Am. **99**, 3572–3584.
Lisowska, G., Smurzynski, J., Morawski, K., Namyslowski, G., and Probst, R. (**2002**). "Influence of contralateral stimulation by two-tone complexes, narrow-band and broad-band noise signals on the 2f1-f2 distortion product otoacoustic emission levels in humans," Acta Oto-Laryngol. **122**, 613–619.
Long, G. R., and Talmadge, C. L. (**1997**). "Spontaneous otoacoustic emission frequency is modulated by heartbeat," J. Acoust. Soc. Am. **102**, 2831–2848.
Long, G. R., Talmadge, C. L., and Lee, J. (**2008**). "Measuring distortion product otoacoustic emissions using continuously sweeping primaries," J. Acoust. Soc. Am. **124**, 1613–1626.
Lutman, M. E., and Deeks, J. (**1999**). "Correspondence amongst microstructure patterns observed in otoacoustic emissions and Bekesy audiometry," Audiology **38**, 263–266.
Maison, S. F., and Liberman, M. C. (**2000**). "Predicting vulnerability to acoustic injury with a noninvasive assay of olivocochlear reflex strength," J. Neurosci. **20**, 4701–4707.
Mott, J. B., Norton, S. J., Neely, S. T., and Warr, W. B. (**1989**). "Changes in spontaneous otoacoustic emissions produced by acoustic stimulation of the contralateral ear," Hear. Res. **38**, 229–242.
Mountain, D. C. (**1980**). "Changes in endolymphatic potential and crossed olivocochlear bundle stimulation alter cochlear mechanics," Science **210**, 71–72.
Müller, J., Janssen, T., Heppelmann, G., and Wagner, W. (**2005**). "Evidence for a bipolar change in distortion product otoacoustic emissions during contralateral acoustic stimulation in humans," J. Acoust. Soc. Am. **118**, 3747–3756.
Murugasu, E., and Russell, I. J. (**1996**). "The effect of efferent stimulation on basilar membrane displacement in the basal turn of the guinea pig cochlea," J. Neurosci. **16**, 325–332.
Purcell, D. W., Butler, B. E., Saunders, T. J., and Allen, P. (**2008**). "Distortion product otoacoustic emission contralateral suppression functions obtained with ramped stimuli," J. Acoust. Soc. Am. **124**, 2133–2148.
Puria, S., Guinan, J. J., Jr., and Liberman, M. C. (**1996**). "Olivocochlear reflex assays: Effects of contralateral sound on compound action potentials versus ear-canal distortion products," J. Acoust. Soc. Am. **99**, 500–507.
Relkin, E. M., Sterns, A., Azeredo, W., Prieve, B. A., and Woods, C. I. (**2005**). "Physiological mechanisms of onset adaptation and contralateral suppression of DPOAEs in the rat," J. Assoc. Res. Otolaryngol. **6**, 119–135.
Reuter, K., and Hammershoi, D. (**2006**). "Distortion product otoacoustic emission fine structure analysis of 50 normal-hearing humans," J. Acoust. Soc. Am. **120**, 270–279.
SAS (**2008**). "SPSS for Mac OS X," SPSS Inc., Chicago, IL.
Shaffer, L. A., and Dhar, S. (**2006**). "DPOAE component estimates and their relationship to hearing thresholds," J. Am. Acad. Audiol. **17**, 279–292.
Shera, C. A. (**2003**). "Mammalian spontaneous otoacoustic emissions are amplitude-stabilized cochlear standing waves," J. Acoust. Soc. Am. **114**, 244–262.
Siegel, J. H., and Kim, D. O. (**1982**). "Efferent neural control of cochlear mechanics? Olivocochlear bundle stimulation affects cochlear biomechanical nonlinearity," Hear. Res. **6**, 171–182.
Stover, L. J., Neely, S. T., and Gorga, M. P. (**1996**). "Latency and multiple sources of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **99**, 1016–1024.
Sun, X. M. (**2008a**). "Contralateral suppression of distortion product otoacoustic emissions and the middle-ear muscle reflex in human ears," Hear. Res. **237**, 66–75.
Sun, X. M. (**2008b**). "Distortion product otoacoustic emission fine structure is responsible for variability of distortion product otoacoustic emission contralateral suppression," J. Acoust. Soc. Am. **123**, 4310–4320.
Talmadge, C. L., Long, G. R., Tubis, A., and Dhar, S. (**1999**). "Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **105**, 275–292.
Wagner, W., Heppelmann, G., Muller, J., Janssen, T., and Zenner, H. P. (**2007**). "Olivocochlear reflex effect on human distortion product otoacoustic emissions is largest at frequencies with distinct fine structure dips," Hear. Res. **223**, 83–92.
Zhang, F., Boettcher, F. A., and Sun, X. M. (**2007**). "Contralateral suppression of distortion product otoacoustic emissions: Effect of the primary frequency in Dpgrams," Int. J. Audiol. **46**, 187–195.
Zwicker, E. (**1990**). "On the frequency separation of simultaneously evoked otoacoustic emissions' consecutive extrema and its relation to cochlear traveling waves," J. Acoust. Soc. Am. **88**, 1639–1641.

Otoacoustic emissions in time-domain solutions of nonlinear non-local cochlear models

Arturo Moleti and Nicolò Paternoster
Dipartimento di Fisica, Università di Roma Tor Vergata, 00133, Rome, Italy

Daniele Bertaccini
Dipartimento di Matematica, Università di Roma Tor Vergata, 00133, Rome, Italy

Renata Sisto and Filippo Sanjust
Dipartimento Igiene del Lavoro, ISPESL, 00040, Monte Porzio Catone, Rome, Italy

(Received 10 June 2009; revised 17 August 2009; accepted 17 August 2009)

A nonlinear and non-local cochlear model has been efficiently solved in the time domain numerically, obtaining the evolution of the transverse displacement of the basilar membrane at each cochlear place. This information allows one to follow the forward and backward propagation of the traveling wave along the basilar membrane, and to evaluate the otoacoustic response from the time evolution of the stapes displacement. The phase/frequency relation of the response can be predicted, as well as the physical delay associated with the response onset time, to evaluate the relation between different cochlear characteristic times as a function of the stimulus level and of the physical parameters of the model. For a nonlinear cochlea, simplistic frequency-domain interpretations of the otoacoustic response phase behavior may give inconsistent results. Time-domain numerical solutions of the underlying nonlinear and non-local full cochlear model using a large number (thousands) of partitions in space and an adaptive mesh in time are rather time and memory consuming. Therefore, in order to be able to use standard personal computers for simulations reliably, the discretized model has been carefully designed to enforce sparsity of the matrices using a multi-iterative approach. Preliminary results concerning the cochlear characteristic delays are also presented. [DOI: 10.1121/1.3224762]

PACS number(s): 43.64.Jb, 43.64.Kc [BLM] Pages: 2425–2436

I. INTRODUCTION

Otoacoustic emissions (OAEs) are a physiological by-product of the activity of the mammalian cochlea (Probst *et al.*, 1991). The OAE generation and backward transmission is effectively described by transmission line cochlear models, including tonotopically resonant transverse impedance terms (e.g., Talmadge *et al.*, 1998). These terms must also model the active feedback mechanism mediated by the outer hair cells (OHCs), which is responsible for the excellent threshold sensitivity and frequency resolution of the mammalian hearing system. A comprehensive cochlear model must be, to some extent, both nonlinear and non-local, and based on the knowledge of the OHC mechanoelectric behavior. Several models of the OHC feedback mechanism have been developed (e.g., Nobili and Mammano, 1996; de Boer and Nuttall, 2003) including detailed analyses of the OHC coupling to the basilar membrane (BM) and to the tectorial membrane, and they have been tested and refined in the past decades through comparison with experimental data, reaching a fairly high degree of complexity, and a correspondingly high number of free parameters. Most models used to predict the OAE generation adopt a simplified view of the OHC active mechanism. This attitude is partly justified by the fact that OAE generation is only a by-product of the cochlear amplifier activity, and the OAE measurable parameters may be critically dependent on cochlear transmission properties other than the details of the local cochlear amplifier at the generation place(s). Nevertheless, some key properties of the OHC physiology must be retained in a full cochlear model, even if one's main purpose is getting correct predictions of the OAE phenomenology.

Nonlinearity is an intrinsic feature of the cochlear physiology, so the frequency-domain solutions of the linearized problem can only approximately predict the behavior of the system, and only in a perturbative regime. Much care must therefore be used when applying to such a system concepts that are fully meaningful for linear systems only, such as the complex frequency response, defined as the Fourier transform (FT) of the impulsive response, or the group delay, defined as the negative slope of the phase/frequency relation. The intrinsically nonlinear equations describing the cochlear micromechanics require, in a nonperturbative regime, a solution in the time domain. On the other hand, the time-domain numerical solutions may become expensive in terms of computational time and memory demanding, if sufficient spatial and time resolutions have to be achieved. High spatial resolution is necessary because the discontinuous variation in the transverse impedance parameters caused by discretization itself must not cause significant spurious reflection of the forward traveling wave (TW). High time resolution is automatically provided by the adaptive integration time step set by the routines used to solve the differential equations, and the related computational cost depends strongly not only on

the number of elements of the discretized cochlea but also on the frequency content of the stimulus and on the characteristic frequencies of the system.

Elliott *et al.* (2007) proposed a matrix formalism, applied to a finite-difference solution method of cochlear models, which is used in this study to model the propagation of the TW and the generation and backward propagation of OAEs. Elliott *et al.* (2007) originally applied this solution scheme to an active linear and local model developed by Neely and Kim (1986). In the Neely and Kim (1986) model, each micromechanical element is a two degree of freedom system of coupled oscillators, simulating some the active cochlear amplifier properties (negative resistance, or anti-damping, in a limited region close to the resonant place). The same scheme can be modified to represent several different cochlear models. In the model by Kim and Xin (2005) (adapted from Lim and Steele, 2002 and generalized to model cochlear impairment in Bertaccini and Fanelli, 2009), the forces applied by the OHCs on the BM are schematized by a nonlinear non-local feed-forward term.

In this work, a feed-forward nonlinear non-local model similar to that proposed by Kim and Xin (2005), in which the OHC additional pressure is assumed proportional to the total pressure on the BM within a slightly more basal region, is implemented in the Elliott *et al.* (2007) semidiscrete scheme, including as well random spatial variations in the impedance parameters (cochlear roughness), which are needed to get appreciable OAE response through coherent reflection (Talmadge *et al.*, 1998), acting as backscattering centers for the forward TW. The semidiscrete model is then fully discretized, and the resulting discrete model is solved efficiently. A nontrivial mass matrix and stiffness of the BM micromechanics (in the numerical analysis sense, i.e., the presence of high Lipschitz constants in the nonlinear model) are suggested using an implicit time-step integrator. Therefore, at each time step, a large system of fully coupled nonlinear algebraic equations should be solved in order to generate the numerical approximations, and this is computationally expensive. In this work, an efficient and reliable numerical simulation is enforced by decoupling the differential part of the discretization of the integrodifferential model by solving sparse linear systems using multi-iterative projection algorithms instead of inverting matrices. A graphical user interface has been added to facilitate the parameters inputted and the analysis of the results. The backward TW associated with OAEs is observed as a displacement wave at the stapes, and some properties of the otoacoustic delays are analyzed.

A variation in the above model was also considered, in which the feed-forward coupling is obtained assuming that the OHC additional pressure is directly proportional to the BM velocity. In this model, this additional force explicitly behaves as a simple anti-damping term.

This discrete model, implemented in our package, after its necessary optimization through comparison with the available data, could be a useful tool to design future OAE experiments, predicting the OAE response at different stimulus levels, and, in particular, to study in more detail the generation place and time of the different components of the OAE response, as well as their direction of propagation along the BM.

In Sec. II, we recall the physical meaning of the cochlear characteristic times that are estimated with different experimental techniques. In Sec. III A we briefly describe the application of Elliott's solution scheme to a very simple one-dimensional (1D) linear, passive, and local cochlear model, to help the reader through Sec. III B, where we discuss its generalization to more realistic, still semidiscrete models, including feed-forward nonlinear and non-local terms. In Sec. III C, we propose a fully discrete feed-forward analog of the underlying model and notes concerning its implementation in the MATLAB environment. In Sec. IV, we discuss our preliminary results, focusing on the relation between different characteristic times in a nonlinear cochlea.

II. BACKGROUND ON COCHLEAR DELAYS IN MODEL AND EXPERIMENT

The study of the characteristic times of the OAE response may provide important information about the cochlear mechanics and the otoacoustic generation mechanisms, complementing other measures coming, e.g., from direct observations of the BM vibration (Ren, 2004; He *et al.*, 2007) and from the analysis of the auditory brainstem response (ABR) (e.g., Neely *et al.*, 1988; Donaldson and Ruth, 1993).

A. TEOAE latency from time-frequency analysis and cochlear transmission delay

In the case of transient evoked OAEs (TEOAEs) and, particularly for click-evoked OAEs (CEOAEs), the latency may be defined in the time domain as the interval between the impulsive stimulus and the onset of the otoacoustic response at a given frequency, which can be measured using time-frequency analysis techniques, based on the wavelet transform or on the MATCHING PURSUIT algorithm (Tognola *et al.*, 1997; Sisto and Moleti, 2002; Jedrzejczak *et al.*, 2004). As the middle ear roundtrip transmission introduces only a very short time delay, of order 0.1–0.2 ms (Puria, 2003; Voss and Shera, 2004), the OAE latency is almost entirely of cochlear origin, being associated for each frequency with the time needed to transmit forward the stimulus from the cochlear base to each tonotopic place as a TW, and backward to the base. This delay is a function of the geometrical and mechanical characteristics of the BM, including those of active filter associated with the feedback mechanism that is mediated by the OHCs. The tonotopic structure of the BM causes the overall decrease in latency with increasing frequency, simply because the cochlear round trip path is longer for lower frequencies. The frequency selectivity of the active cochlear filter is also related to the OAE delay, which increases by increasing the tuning factor Q of the resonance. As a consequence, the OAE latency is a decreasing function of both frequency and stimulus level (Sisto and Moleti, 2007). This property has also been found in the wave-V ABR delay, which is made up (Eggermont and Don, 1980; Neely *et al.*, 1988; Don *et al.*, 1993; Donaldson and Ruth, 1993;

Abdala and Folsom, 1995) of a constant term of neural origin (this is surely true for its main contribution, the delay between wave-I and wave-V), independent of frequency and stimulus level, and of a cochlear term, decreasing with increasing frequency and stimulus level, which is evidently associated with the forward cochlear transmission delay.

Transmission line cochlear models (Furst and Lapid, 1988; Talmadge *et al.*, 1998; Shera *et al.*, 2005) are usually in agreement in representing the acoustic signal propagation along the BM as a TW. Due to the tonotopicity of the BM, each Fourier component of the stimulus propagates up to its resonant place, where it produces the maximum transversal displacement of the BM, associated with that tone perception, and then it is locally absorbed.

B. OAE generation mechanisms

It is generally accepted that OAEs are produced by two different mechanisms: nonlinear distortion and linear reflection (Shera and Guinan, 1999).

The cochlear response nonlinearity generates distortion at moderate and high BM excitation levels. A threshold for the onset of nonlinearity can be fixed at some transverse displacement amplitude of order 10 nm. At these stimulus levels, significant OAE generation is expected from the nonlinear generation mechanism. If a given cochlear region is simultaneously excited by two tones of different frequencies, the system nonlinearity also produces tones at frequencies that are linear combinations of those of the stimulus, as in the case of the distortion product OAEs (DPOAEs). The nonlinear distortion generation always occurs at the tonotopic cochlear place of the stimulus frequency, or, as in the DPOAE case, in a place that is a well-defined function of the frequencies of the stimulus (primary tones); this generation mechanism is therefore called "wave-fixed" (Shera and Guinan, 1999). Linearized transmission line cochlear models (Talmadge *et al.*, 1998, 2000; Shera *et al.*, 2005), solved in the frequency domain, predict for OAEs generated by wave-fixed mechanisms a flat phase spectrum (at least, in the scale-invariant limit). If the resulting null "group delay" were simplistically interpreted as instantaneous cochlear response, there would be obvious contradiction with the hypothesis that the otoacoustic response is generated at (or near) the tonotopic place, for each frequency, because at least the forward propagation of the stimulus (one could argue that the backward OAE propagation could be much faster) would need a significant and well-measurable transmission time, from a few to several milliseconds, dependent on frequency.

OAE generation is also expected to be associated with the reflection of a significant fraction of the forward TW. It is necessary to postulate the presence of randomly distributed microirregularities of the cochlear mechanical structure, which act as backscattering centers for the forward TW (Zweig and Shera, 1995). Perturbative estimates of the cochlear reflectivity, based on the osculating parameters technique (Shera and Zweig, 1991; Talmadge *et al.*, 2000), suggest that most of the linearly reflected wave should come from a region close to the resonant place. Recent estimates based on a linear model by Choi *et al.* (2008) suggest instead that a significant contribution to the overall SFOAE response may come from cochlear regions remote from the resonant place. Coherent reflection from a rather broad cochlear region slightly basal to the resonant place is predicted by the coherent reflection filtering (CRF) theory, which also introduces time-delayed stiffness terms in the BM micromechanics to get the necessary tall and broad activity pattern. These "Zweig" terms, due to their rather fine-tuned delays, act as effective damping and anti-damping terms in that cochlear region, providing the negative resistance region required by the solution of the inverse cochlear problem applied to experimental BM transfer function data. As already remarked, similar results could be obtained with different mathematical approaches, e.g. modeling each cochlear partition as a two degree of freedom system of linear actively coupled oscillators (Neely and Kim, 1986), or by introducing active nonlocal terms. In the CRF theory, the OAE generation mechanism is considered "place-fixed" because the backscattering centers are localized at fixed positions. The CRF theory predicts, for such a place-fixed mechanism, a rapidly rotating phase spectrum (Talmadge *et al.*, 1998; Shera *et al.*, 2005). The additional reflection from more basal cochlear regions suggested by Choi *et al.* (2008) would imply a flatter phase-frequency relation, and the vector superposition of the two sources would explain the observed stimulus-frequency OAE (SFOAE) spectral fine structure, without having to assume contributions from nonlinear distortion.

C. Relation among different cochlear characteristic times

Much experimental evidence has been gathered about the relation among different cochlear characteristic times. A general warning applies to such comparisons. In classical BM transfer function measurements (e.g., Rhode, 1971), the place of measurement is fixed as the frequency changes. The phase slope represents a partial derivative. In OAE experiments, this is not generally the case. The relation between OAE phase-gradient delays and the actual time delay of each frequency component of the OAE response is not straightforward, and depends on the wave-fixed or place-fixed nature of the OAE generator. OAE phase-gradient delays are actually measured by computing the slope of the phase-frequency relation, but the phase is a function of both the frequency of the OAE and the position of the source. The experimentally measured slope is therefore a total derivative, which includes an additional term for wave-fixed generation, which almost totally cancels, in the WKB approximation (Sisto *et al.*, 2007), the one associated with the roundtrip transmission delay. For place-fixed generation, instead, the phase is a function of frequency only, and the phase-gradient delay is therefore expected to approximately coincide with the physical transmission delay.

At low stimulus levels, the OAE response should be dominated by linear place-fixed mechanisms, the SFOAE phase-gradient delay is therefore expected to be closely related to the physical delay associated with the forward and backward transmissions along the BM. Early applications of the CRF theory predicted indeed that the SFOAE phase-gradient delay is twice the BM group delay (Shera and

Guinan, 2003), while refined calculations (Shera *et al.*, 2005) concluded that the phase-gradient delay should be slightly less than twice the BM group delay.

Accurate studies on chinchillas (Siegel *et al.*, 2005) demonstrated that the SFOAE phase-gradient delay is significantly less than twice the BM group delay, particularly at low-frequency. This result has been considered in agreement with the CRF theory assuming that some contribution from nonlinear distortion is also present in the SFOAE response (Shera *et al.*, 2006). Another explanation of these discrepancies could be provided by the linear reflection sources remote from the resonance place proposed by Choi *et al.* (2008). It has also been shown (Sisto *et al.*, 2007) that the roundtrip cochlear delay measured by time-frequency analysis of CEOAE waveforms closely matches the phase-gradient delay of the same OAE spectra, at click stimulus levels from 60 to 90 dB peak SPL (pSPL), concluding that linear reflection from roughness is the main source of TEOAEs in this stimulus level range. We recall that the pSPL level of a click is given by the ratio between its peak amplitude and the standard reference pressure level (20 μPa), expressed in decibels. The associated spectral density is a function of the click duration.

For DPOAEs, the relation between latency, phase-gradient delay, BM group delay and frequency is even less straightforward, and depends on the experimental sweeping paradigm (Prijs *et al.*, 2000; Schoonhoven *et al.*, 2001). If the ratio f_2/f_1 is kept constant, from the solution of the linearized cochlear equations (Talmadge *et al.*, 2000), it follows that the nonlinear distortion component, originated in $x(f_2)$, should have flat phase spectrum, while the linear reflection source coming from $x(f_{DP})$ should give a contribution with rapidly rotating phase. In the time domain, it is clear that the nonlinear generation may start only after the transmission time needed for the f_1 and f_2 tones to reach the nonlinear generation place $x(f_2)$. After that, an additional (shorter) time is necessary for the backward TW at frequency f_{DP} to reach the base. The backward delay is shorter because the wave packet of frequency f_{DP} moves away from a region that is already more basal than its tonotopic region, where its group velocity would be considerably lower (Moleti and Sisto, 2003) (this slowing-down effect near the resonant place is sometimes called "filter build-up time," it may be seen as a significant contribution to the path integral of the inverse velocity coming from a rather short part of the path). The second DPOAE source [from linear reflection at $x(f_{DP})$] is significantly more delayed because the distortion tone generated in $x(f_2)$ must reach its own tonotopic place $x(f_{DP})$ to be amplified and reflected back by roughness. Therefore, its overall onset delay in the time domain is expected to be close to the latency of the component of frequency f_{DP} of a correspondently high TEOAE response [neglecting the dependence on the stimulus level, which makes a little shorter the forward transmission delay to $x(f_2)$ of the primary tones f_1 and f_2, due to their higher level]. Summarizing, observing the phase of the two DPOAE components in the frequency domain, one should see a flat phase component and a rotating phase component (which can be separated using time-domain filtering), whereas observing the onset of the response in the time domain, one should observe an early (but not instantaneous) response from the first source and a more delayed onset of the linear reflection contribution. Long *et al.* (2008) recently exploited the different phase behaviors of the two DPOAE components to separate them by using sufficiently fast-sweeping primary tones. Whitehead *et al.* (1996) were able to measure (in humans and rabbits) the onset time of DPOAEs elicited by high-level primaries [75 dB sound pressure level (SPL)] using a clever differential acquisition technique based on phase rotation of the primary tones. They measured delays from 2 to 5–10 ms (depending also on the data analysis algorithm) in the 1–8 kHz range, with the expected frequency dependence. These delays are compatible with those expected from the nonlinear distortion source, which should be dominant at high stimulus levels.

D. OAE backward propagation

By comparing the OAE latency to the ABR wave-V latency in the same stimulus level range, it has been shown that the part of the ABR latency that is independent of frequency and stimulus level (associated with the forward cochlear path of the stimulus) is approximately equal to half the OAE latency (Moleti and Sisto, 2008), supporting the hypothesis that the backward propagation of OAEs is due to a slow transverse TW on the BM. This conclusion is in agreement with analyses of the data from Allen–Fahey experiments (Shera *et al.*, 2007), and with direct measurements by Dong and Olson (2008), but it is contradicted by the slope of the DP phase at different cochlear places measured by accurate observations of the BM vibration either by moving the observation place (He *et al.*, 2007) or by moving the primary frequencies (de Boer *et al.*, 2008). These contradictory observations could be attributed to a dominant forward traveling DP within the cochlea, which would obscure the observation of reverse waves.

The above list of interesting issues, which can only be approximately evaluated with frequency-domain solutions, due to the intrinsic nonlinearity of the problem, was meant to demonstrate the strong need for time-domain solutions of the full cochlear problem. In the following, we will discuss some preliminary results from the time-domain solution of a nonlinear non-local active cochlear model, focusing on possible applications to the study of the OAE delays.

III. COCHLEAR MODELS

A. Linear 1D box model

In this subsection, we apply the scheme of Elliott *et al.* (2007) to a very simple linear passive model, to help the reader getting through the formalism before going to Sec. III B, where the feed-forward model is described. A list of the parameter values used in the model is reported in Table I.

For an incompressible fluid, in a cochlear duct of rectangular constant cross section of constant half-height H and length L, divided by a tonotopically resonant elastic BM, the wave propagation along the cochlea on the BM ($z=0$), reduces to the 1D transmission line equation for the differential pressure p:

TABLE I. Model parameters used in this study. Some of the parameter values listed below are taken from Talmadge *et al.* (1998), they are indicated by (T98).

Parameter	Value	Definition
ρ	10^3 kg m^{-3}	Fluid density
L	3.5×10^{-2} m	Length of the BM
k_0	3.1×10^{3} m^{-1}	Cochlear geometrical wavenumber (T98)
ω_0	$2.08\times10^{4}\cdot2\pi$ s^{-1}	Greenwood's map frequency coefficient (T98)
σ_{bm}	5.5×10^{-2} kg m^{-2}	BM density (T98)
ω_1	$-145\cdot2\pi$ s^{-1}	Greenwood's map frequency offset (T98)
k_ω	1.382×10^{2} m^{-1}	Greenwood's map inverse length scale (T98)
γ_0	5.035×10^{3} s^{-1}	Cochlear damping map coefficient (T98)
γ_1	100 s^{-1}	Cochlear damping map offset (T98)
k_Γ	1.382×10^{2} m^{-1}	Cochlear damping map inverse length scale (T98)
K_{ow}	2×10^{8} N m^{-3}	Effective middle ear-oval window stiffness
γ_{ow}	5×10^{3} s^{-1}	Effective middle ear-oval window damping
σ_{ow}	2 kg m^{-2}	Effective middle ear-oval window density
ξ_{nl}	10^{-8} m	OHC gain saturation length scale
γ	0.36	OHC gain parameter
λ	1.2×10^{-7} m^2	OHC non-local interaction range (squared)

$$\frac{\partial^2 p(x,0,t)}{\partial x^2} = \frac{2\rho}{H}\ddot{\xi}(x,t), \tag{1}$$

where ρ is the fluid density and ξ is the BM transverse displacement at the longitudinal position x and time t.

Equation (1) is obtained, as usual, assuming the fluid incompressibility, using the boundary condition on the BM:

$$\left.\frac{\partial p(x,z,t)}{\partial z}\right|_{z=0} = 2\rho\ddot{\xi}(x,t), \tag{2}$$

and that on the rigid upper wall:

$$\left.\frac{\partial p(x,z,t)}{\partial z}\right|_{z=H} = 0. \tag{3}$$

As in Elliott *et al.*, 2007, the first of the N elements of the semidiscretized model includes the middle ear dynamics and the boundary condition for the wave equation (1) at the basal end:

$$\left.\frac{\partial p}{\partial x}\right|_{x=0} = 2\rho\ddot{\xi}_{ow}, \tag{4}$$

where $\ddot{\xi}_{ow}$ is the acceleration of the stapes. Elliott *et al.* (2007) wrote this acceleration as the linear combination of two components: the acceleration due to external excitation and that due to the loading by the internal pressure response in the cochlea at $x=0$. We prefer to choose a slightly different approach, putting the term associated with the stimulus in the ear canal as a forcing term in the dynamical equation for the first element of the partition, according to Eq. (10) of Talmadge *et al.*, 1998:

$$\ddot{\xi}_{ow}(t) + \gamma_{ow}\dot{\xi}_{ow}(t) + \omega_{ow}^2\xi(t) = \frac{p(0,t) + G_{me}P_{dr}(t)}{\sigma_{ow}}, \tag{5}$$

where, γ_{ow}, $K_{ow}=\omega_{ow}^2\sigma_{ow}$, and σ_{ow} are the phenomenological parameters reported in Table I, chosen to represent the filtering properties of the middle ear, P_{dr} is the calibrated pressure in the ear canal (for a rigid ear drum), and G_{me} is the middle ear mechanical gain of the ossicles.

The last element of the spatially discretized cochlea is the helicotrema, which is described, as usual, by a pressure release (short-circuit) boundary condition:

$$p(L,z,t) = 0. \tag{6}$$

Considering the dynamical equation that relates the BM transversal displacement to the p acting on the tonotopic oscillator, we have for the elements from 2 to $N-1$:

$$\ddot{\xi}(x,t) + \gamma_{bm}(x,\xi,\dot{\xi})\dot{\xi}(x,t) + \omega_{bm}^2(x,\xi,\dot{\xi})\xi(x,t) = \frac{p(x,0,t)}{\sigma_{bm}}. \tag{7}$$

The height H is related to the BM density and to the cochlear geometrical wavenumber k_0, defined in Talmadge *et al.*, 1998 by $H=2\rho/k_0^2\sigma_{bm}$.

In the simplest form of the model, each tonotopic place is schematized by a single passive oscillator, and both damping and stiffness are smooth functions of the x only, according to the Greenwood map (Greenwood, 1990):

$$\omega_{bm}(x) = \omega_0 e^{-k_\omega x} + \omega_1,$$

$$\gamma_{bm}(x) = \gamma_0 e^{-k_\Gamma x} + \gamma_1. \tag{8}$$

In the limit $k_\Gamma=k_\omega$, and $\omega_1=\gamma_1=0$, the map is also scale-invariant. This symmetry is violated in the real cochlea, particularly at low-frequency, due to the constant terms ω_1 and γ_1, and also because $k_\Gamma\neq k_\omega$. Indeed, cochlear tuning $Q=\omega/\gamma$ increases with frequency, as shown by behavioral and otoacoustic data (e.g., Glasberg and Moore, 1990; Shera *et al.*, 2002; Unoki *et al.*, 2007; Sisto and Moleti, 2007).

Elliott *et al.* (2007) described each element of the cochlear partition as a system of two coupled linear oscillators, including active terms, according to a model by Neely and Kim (1986). In this study, we chose a different approach, describing each partition with a single oscillator, and introducing, in the next subsection, active amplification and nonlinear saturation terms as additional forces triggered by the OHCs and acting on the BM, generated by non-local feedforward longitudinal interaction, similar to what has been proposed by Kim and Xin (2005), see also Bertaccini and Fanelli, 2009, in different solution schemes.

Using finite-difference approximation for the spatial derivatives, the semidiscrete models can be written in matrix form

$$FP(t) = \ddot{\Xi}(t), \tag{9}$$

where F is Elliott's $N\times N$ finite-difference matrix, whose first and last lines include, respectively, the boundary conditions, Eqs. (4) and (6), $P(t)$ and $\ddot{\Xi}(t)$ are the N-dimensional

vectors of the differential pressure and cochlear partition acceleration, respectively:

$$F = \frac{H}{2\rho(\Delta x)^2}\begin{bmatrix} -\frac{\Delta x}{H} & \frac{\Delta x}{H} & 0 & & & & 0 \\ 1 & -2 & 1 & 0 & & & 0 \\ 0 & 1 & -2 & 1 & & & 0 \\ & & \cdots & 1 & -2 & 1 & 0 \\ & & & & 1 & -2 & 1 \\ 0 & & & & & 0 & -\frac{2\rho(\Delta x)^2}{H} \end{bmatrix}, \tag{10}$$

where $\Delta x = x_i - x_{i-1} = L/(N-3)$.

As in Elliott *et al.*, 2007, we cast the dynamic variables $[\dot{\xi}_j(x_j,t),\ \xi_j(x_j,t)]$ of the micromechanical elements in a single vector of state variables U of dimension $2N$.

Equations (4)–(7) can be written for the whole set of discrete tonotopic oscillators in the form of combined matrix equations:

$$\dot{U}(t) = A_E U(t) + B_E(P(t) + S(t)), \tag{11a}$$

$$\dot{\Xi}(t) = C_E U(t), \tag{11b}$$

where $S(t)$ is a vector whose only non-null element is the first one, which is equal to $G_{me}P_{dr}(t)$.

The matrices A_E $(2N\times 2N)$, B_E $(2N\times N)$, and C_E $(N\times 2N)$ are block diagonal. In particular, each block A_i of the matrix A_E, for $i=2,\ldots,N-1$, contains the dynamics of the ith resonant tonotopic oscillator:

$$A_E = \begin{bmatrix} A_1 & & \\ & \cdots & \\ & & A_N \end{bmatrix} \text{ (the same rule applies to } B_E \text{ and } C_E), \tag{12}$$

with

$$A_i \equiv \begin{bmatrix} -\gamma_{bm}(x_i) & -\omega_{bm}^2(x_i) \\ 1 & 0 \end{bmatrix}, \quad B_i \equiv \begin{bmatrix} \frac{1}{\sigma_{bm}} & 0 \end{bmatrix}^T \quad \text{for} \quad i = 2,\ldots,N-1,$$

$$A_1 \equiv \begin{bmatrix} -\gamma_{ow} & -\omega_{ow}^2 \\ 1 & 0 \end{bmatrix}, \quad B_1 \equiv \begin{bmatrix} \frac{1}{\sigma_{ow}} & 0 \end{bmatrix}^T, \quad A_N \equiv 0, \quad B_N \equiv 0,$$

$$C_i \equiv [1 \quad 0]. \tag{13}$$

The finite-difference matrix F is invertible, so we can write Eq. (9) as

$$P(t) = F^{-1}\ddot{\Xi}(t) = F^{-1}C_E\dot{U}(t). \tag{14}$$

Substituting Eq. (14) into Eq. (11a), the overall state space equation with distributed micromechanics and boundary conditions can be written in the general form

$$M_{lin}\dot{U}(t) = A_E U(t) + B_E S(t), \tag{15}$$

where M_{lin} is the $2N\times 2N$ mass matrix of the system:

$$M_{lin} = I - B_E F^{-1} C_E. \tag{16}$$

B. Nonlinear feed-forward model

In a more advanced model, the OHCs-BM interaction can be schematized as a nonlinear, non-local active system that can be included into the same matrix solution scheme.

In the model by Kim and Xin (2005), the pressure applied by OHCs on the BM is assumed proportional to the total pressure on the BM, and, due to the longitudinal tilt of OHCs, forces acting on the cilia at x cause OHCs to push at a point $x+\Delta$ downstream on the BM:

$$q(x+\Delta,t) = \alpha(\xi,x,t)p_{\mathrm{BM}} = \alpha(\xi,x,t)(p(x,t) + q(x,t)), \tag{17}$$

where q is the additional pressure given by the OHCs, p_{BM} is the total pressure on the BM, and α is a nonlinear non-local gain factor, which depends on the BM displacement ξ in a cochlear region around the considered position x.

For the gain function, we use the integral expression (Kim and Xin, 2005):

$$\alpha(x,\xi,t) = \frac{\gamma}{\sqrt{\lambda\pi}}\int_0^L \exp\left(-\frac{(x-x')^2}{\lambda}\right) g(\xi(x',t))dx', \tag{18}$$

where γ is a dimensionless parameter controlling the strength of the non-local terms, and $\sqrt{\lambda}$ is a characteristic length (a constant in a scale-invariant cochlea), representing the longitudinal range of the non-local interaction.

Here we choose the nonlinear analytical gain function $g(\xi(x,t))$:

$$g(\xi(x,t)) = \tanh\left(\frac{\xi_{nl}^2}{(\xi(x,t) - \xi_0)^2}\right), \tag{19}$$

which approximately matches the nonlinear gain function shown by Kim and Xin (2005) and by Lim and Steele (2002), where ξ_{nl} is a transverse BM displacement scale for the nonlinear saturation of the OHC gain, and ξ_0 is a parameter controlling the vertical asymmetry of the OHC gain (in our simulations $\xi_{nl}=10^{-8}$ m and $\xi_0=0$). This is surely an oversimplified version of the actual physiology of the OHC mechanism, which is much more accurately described elsewhere (e.g., Nobili and Mammano, 1996). The inclusion of a more realistic description of the OHC physiology, which would increase the complexity of the numerical solution of the problem and would also introduce a much higher number of parameters, is beyond the scope of the present study.

Including Eq. (17), Eq. (7) is modified as follows:

$$p(x,t) + q(x,t) = \ddot{\xi}(x,t) + \gamma_{bm}(x)\dot{\xi}(x,t) + \omega_{bm}^2(x)\dot{\xi}(x,t),$$

$$q(x,t) = \alpha(x-\Delta,\xi,t)(p(x-\Delta,t) + q(x-\Delta,t)) \quad (\text{for } \Delta \le x \le L),$$

$$q(x,t) = 0 \quad (\text{for } 0 \le x \le \Delta). \tag{20}$$

In the semidiscrete model, the feed-forward term is related to the pressure by

$$q(x_i,t) - \alpha(x_{i-K},\xi,t)q(x_{i-K},t) = \alpha(x_{i-K},\xi,t)p(x_{i-K},t), \tag{21}$$

where K is an integer number such that

$$\Delta = K\Delta x. \tag{22}$$

Equation (24) can be expressed as a matrix equation:

$$BQ(t) = CP(t). \tag{23}$$

$Q(t)$ and $P(t)$ are, respectively, the column vectors for $q(x_i,t)$ and $p(x_i,t)$. The matrix B has 1's on its diagonal and off-diagonal nonzero elements:

$$B(i+K,i) = -\alpha(x_i,\xi,t) \quad \text{for} \quad i = 2,\ldots,N-K. \tag{24}$$

The matrix C is a matrix whose nonzero elements are

$$C(i+K,i) = \alpha(x_i,\xi,t) \quad \text{for} \quad i = 2,\ldots,N-K. \tag{25}$$

The B and C matrices are both functions of the BM displacement. In particular, B is invertible.

After same manipulations, it can be shown that the following equation for the state vectors U holds

$$M_{nl}\dot{U}(t) = A_E U(t) + B_E S(t), \tag{26}$$

where the nonlinear mass matrix is

$$M_{nl} = (I - B_E G(U) F^{-1} C_E), \tag{27}$$

and $G(U)$ is the $N \times N$ gain matrix:

$$G(U) = B^{-1}C + I. \tag{28}$$

In the limit in which the nonlinear coupling term α is zero, the matrix B reduces to the identity matrix and C is zero. In this limit the gain matrix, $G(U)$ is coincident with the identity matrix. This is the limit in which the linear passive equations hold, Eq. (26) reduces to Eq. (15), and the mass matrix of the system reduces to M_{lin} [Eq. (16)]. For $\alpha \neq 0$, in the case $K=0$, there is no feed-forward asymmetry, but the model is still nonlinear and non-local. Different values of K could be chosen, providing the desired amount of asymmetry, to match the experimentally measured shape of the BM activity patterns.

The same scheme could be easily adapted to describe different models of the OHC function. For example, one could assume that the additional OHC pressure is proportional to the BM velocity. This assumption may be questionable on a physiological basis, but it is, however, interesting to note that it would lead to a model in which the OHC force would act as an explicit anti-damping term everywhere along the BM, not only near the resonant place. As this assumption is implicitly made, when one uses simple 1D transmission line models in which the anti-damping term is just a negative damping constant at each cochlear place and saturation is given by a quadratic damping term (e.g., a Van der Pol oscillator model), it could be interesting to compare the time behavior of OAEs produced by such a model with that predicted by the previous one.

Equation (21) would formally change to

$$q(x_i,t) = \alpha'(x_{i-K},t)\sigma_{bm}\gamma_{bm}(x_i)\dot{\xi}(x_{i-K},t), \tag{29}$$

where $\gamma(x_i)$ is the local damping constant and α' is obtained from an integral like that of Eq. (18), with a different value γ' of the dimensionless constant γ that controls the stability of the resonance.

The mass matrix of Eq. (27) becomes

$$M'_{nl} = I - B_E(F^{-1}C_E + CD_E), \tag{30}$$

where D_E is a block diagonal matrix, whose elements are

$$D_{i>1} = [0 \;\; \sigma_{bm}\gamma_{bm}(x_i)], \quad D_1 = [0 \;\; 0], \tag{31}$$

whereas the other matrices are unchanged.

If the nonlinear gain function $g(\xi(x,t))$ is also changed to

$$g'(\xi(x,t)) = 1 - \frac{\xi^2}{\xi_{nl}^2}, \tag{32}$$

one gets a nonlinear non-local model with explicit anti-damping and quadratic nonlinear damping at each cochlear place x:

$$\gamma'(x,\xi) = \gamma_{bm}(x)\left(-\alpha' + 1 + \alpha'\frac{\xi^2}{\xi_{nl}^2}\right). \tag{33}$$

We note that the generalization to a wide class of different models is a simple task, in the scheme of Elliott *et al.* (2007), exploiting the fact that one is free to select the BM velocity as the first component of each element of the state vector $\dot{U}$ or as the second component of each element of U using the matrices C_E and D_E, respectively. This freedom of choice is important because it allows one to put the nonlinear term into the mass matrix of the system.

C. A fully discrete nonlinear active model and its numerical implementation

In this section, we discuss a numerical approximation technique for the *semidiscrete* model (26). We recall that semidiscrete means that we have to do it with a model that is no longer based on partial differential equations, still continuous, but now only ordinary derivatives are present. We consider a uniform mesh on a rectified model of the BM. Discretization with respect to the spatial variable x imposed on the BM gives the sequence of systems of nonlinear integrodifferential equations (26), (18), and (19) with null initial conditions, where each of the systems as in Eq. (26) is parametrized by the spatial step Δx of the mesh. We recall that the integral part of both the continuous and the semidiscrete model (26) is due to the nonlocality of the gain factor $\alpha(\xi,x,t)$ in Eq. (18) and hidden in the matrix functions B, C, and $G(U)$ in Eq. (26), computed for each time step. In order to simplify the overall calculation and to avoid potential instabilities, we computed the gain factor by using information from the previous step; i.e., we considered a sort of semi-implicit reduction of the models (26), (18), and (19).

We remark that the differential systems in Eq. (26) have a nontrivial mass matrix whose expression can be simplified in

$$M_{nl} = I - B_E B^{-1} F^{-1} C_E, \tag{34}$$

by using Eq. (28) and observing that $C=-B+I$; i.e., the expression of the gain matrix can be reduced to $G(U)=B^{-1}$.

In order to provide time-step integration of Eq. (26), we observe that, whenever the mass matrix is different from the identity, using a package based on implicit or on explicit formulas has similar computational costs. In particular, in order to advance in time, one needs to solve algebraic nonlinear equations requiring the solution of algebraic linear systems of *neq* equations, where *neq* is the number of single differential equations in Eq. (26) even using a code based on explicit formulas. Therefore, we modified for the use of a multi-iterative procedure a stable package based on Backward Differentiation Formulas (BDF-like) variable step, variable order (from order 1 to 5) formulas that are of implicit type, ode15s, that is part of MATLAB, by Mathworks©. The underlying linear algebraic systems to be solved at each time step of ode15s have matrices that can be decomposed in the form

$$A = M_{nl} - \delta t \cdot a \cdot J, \tag{35}$$

where M_{nl} is the mass matrix, δt is the actual time step, J is the Jacobian matrix, and a is a constant. In our setting, the Jacobian matrix J is constant. On the other hand, we stress that the mass matrix M_{nl} is not and does depend on the solution, i.e., on the BM position ξ. Moreover, the formal expression of M_{nl} in Eqs. (26), (30), and (35) includes the inverse of matrix function B (lower bidiagonal, changing with the solution ξ at each time step) and of the matrix F (tridiagonal, constant, generated by the five-point finite-difference discretization of the Laplacian) that are full matrices; i.e., all entries of B^{-1}, F^{-1} are different from zero. Therefore, in order to avoid full coupling of the differential systems (26), requiring a computational cost per time step of $O((2N)^3)$ and a storage for $O(N^2)$ double precision floating point entries, we should not invert any matrix explicitly. Unfortunately, due to the nature of the matrix A in Eq. (35) as a sum of two components, we cannot use direct solvers for the linear systems of the form $Ax=b$. A popular way to approach this is the use of fixed-point iteration algorithms, as in Kim and Xin, 2005. However, fixed-point iteration algorithms converge slowly and often impose restrictions on the parameters of the model for convergence. In particular, artificial restrictions on the time step and/or on the admitted values of some parameters are an issue and this was the case for Kim and Xin (2005) approach, see Bertaccini and Fanelli, 2009 for a way to overcome this. In view of this, we propose here the use of iterative Krylov subspace solvers as the core solver for the linear algebraic systems with matrices as in Eq. (35). Indeed, by using iterative Krylov subspace solvers, we are able to lower the computational cost per step to at most linear in N (the number of mesh points on the BM). We recall that working with iterative Krylov subspace solvers does not require forming or storing the matrix A or its intermediate components. It is enough to access A through matrix-vector products, e.g,, there is no need to form M_{nl}. The only requirement is to provide a fast procedure that, given a vector v, computes the vector

$$w = A \cdot v, \tag{36}$$

and the operation (36) is performed at each iteration of the Krylov solvers. Therefore, the operation (36) becomes the core "outer" operation in the solution of the underlying discretized model and it is performed through the solution of several "inner" steps, consisting in the solution of two sparse linear systems by sparse direct methods and in some other matrix-vector products and linear combinations of vectors. We experienced that the Krylov subspace iterative solver chosen, GMRES, converges to the required tolerance within a moderate average number of iterations that does not increase with N, the mesh size on the BM. More details on the technical solutions adopted and an analysis of the convergence process will be given in a forthcoming paper.

IV. RESULTS

In this section, we present some preliminary results to show that the nonlinear and non-local model described in Sec. II B, fully discretized and optimized in Sec. II C, may become, after having been tuned by careful comparison with the experimental available data, a useful complement to future experiments, to study some of the OAE issues mentioned in Sec. III. In the following, we show that the model is able to produce OAEs as a response to both impulsive (TEOAEs) and stationary stimuli (DPOAEs). The simulated TEOAEs show the expected time-frequency behavior, with shorter latency at higher frequency, consistent with the hypothesis that their backward transmission is associated with a slow transverse TW on the BM. The DPOAE components are produced at the cochlear places predicted by the theory after the corresponding forward transmission delays.

A. TEOAEs

The result of a numerical simulation ($N=1000$ partitions) using a broad-band click stimulus (level corresponding to 80 dB pSPL, duration of 80 μs in the ear canal, similar to that routinely used in the clinical practice) is shown in Fig. 1(a), where we plot the computed BM transverse displacement as a function of time (in a 20 ms interval) and cochlear position x. One could choose to plot the BM velocity instead of displacement, obtaining a different vertical shape, due to the factor ω, which would amplify the basal part of the TW. From the top view shown in Fig. 1(b), the expected relation between the forward transmission time delay (BM forward delay) and the position $x(\omega)$ of the tonotopic resonant place of each frequency component ω is more clearly visible. This relation may be converted into a relation between BM delay and frequency using the Greenwood map (Greenwood, 1990).

Including randomly distributed mechanical irregularities (roughness) as spatial stiffness variations in relative amplitude $\varepsilon=0.05$, the click stimulus produces the delayed response at the stapes shown (for a total time of 50 ms) in Fig. 2 (the data are windowed to cancel the stimulus and to allow spectral analysis). This response would be transmitted back through the middle ear producing a TEOAE in the ear canal. A high level of fluctuations is used in the example to get a

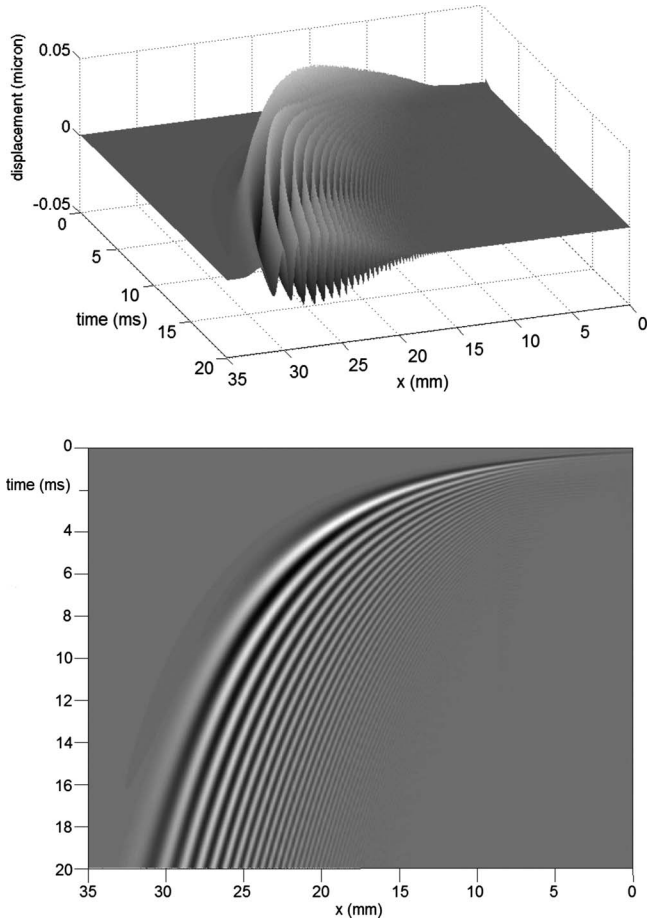


FIG. 1. BM response to a broad-band pulse (an 80 dB pSPL click of duration 80 μs), as a function of time and cochlear longitudinal position x.

strong TEOAE signal. The waveform of Fig. 2 has been analyzed using time-frequency wavelet techniques to estimate the time delay of each frequency component. This delay closely corresponds to that of the TEOAE that would be measured in the ear canal because the delay introduced by the middle ear transmission, neglected in this model, is negligible (of order 100–200 μs). The TEOAE wavelet delay

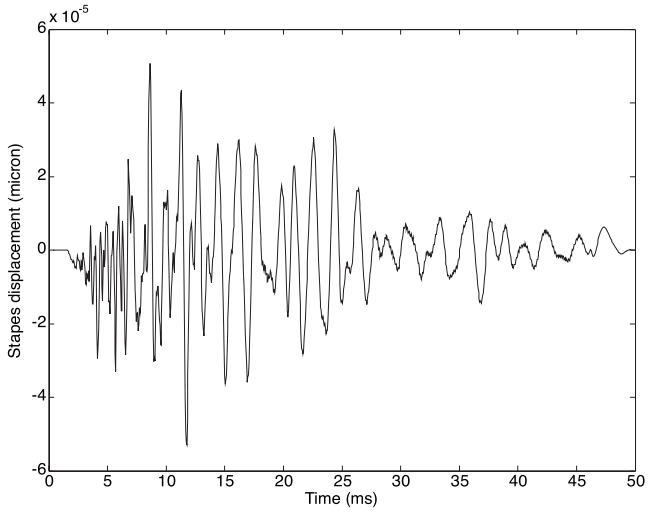


FIG. 2. "Otoacoustic" response computed at the stapes for 50 ms after the click, corresponding to the cochlear activation pattern of Fig. 1.

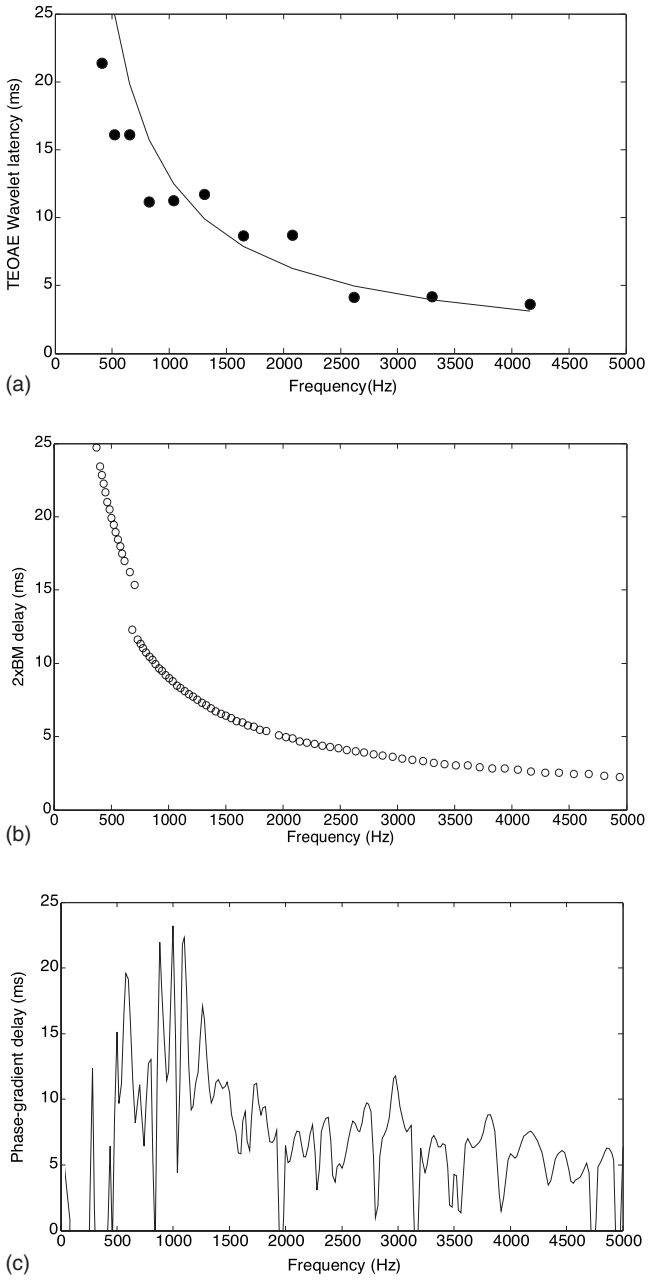


FIG. 3. Wavelet analysis estimate of the latency/frequency relation (a) of the response at the base of Fig. 2, compared to twice the delay of the BM response at each tonotopic place (b) and to the phase-gradient delay measured from the FFT of the same waveform (c).

computed for the waveform of Fig. 2 is shown in Fig. 3(a). In Fig. 3(b), we show twice the BM forward latency, estimated from Fig. 1(b) as the time of the maximum BM excitation and attributed to the frequency that is the best frequency for each place according to the Greenwood map (Greenwood, 1990). In Fig. 3(c), we show the phase-gradient delay estimated from the slope of the fast Fourier transform (FFT) phase. The good agreement confirms that the signal observed at the stapes comes from a backward slow TW on the BM, generated, for each frequency component of the stimulus, near its resonant place. In this model, the backward wave is generated by linear reflection from roughness. Indeed, the same simulation without roughness (not shown) produces no "OAE" response at the stapes. We note that the

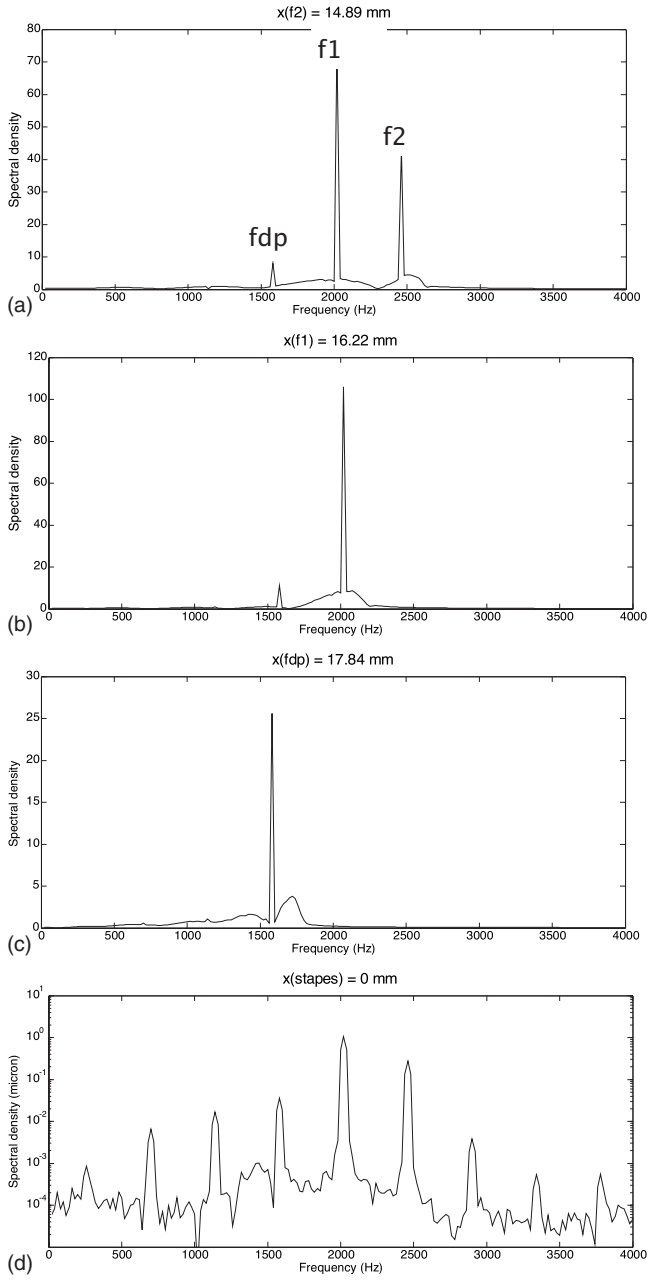


FIG. 4. Generation of the $2f_2-f_1$ distortion product due to nonlinear interaction of two primary tones ($f_1=2000$ Hz, $f_2/f_1=1.22$). The distortion tone is generated at $x(f_2)$(a), its amplitude constantly increases reaching first $x(f_1)$ (b), and then $x(f_{DP})$ (c). The response at the stapes includes several other DP lines (d).

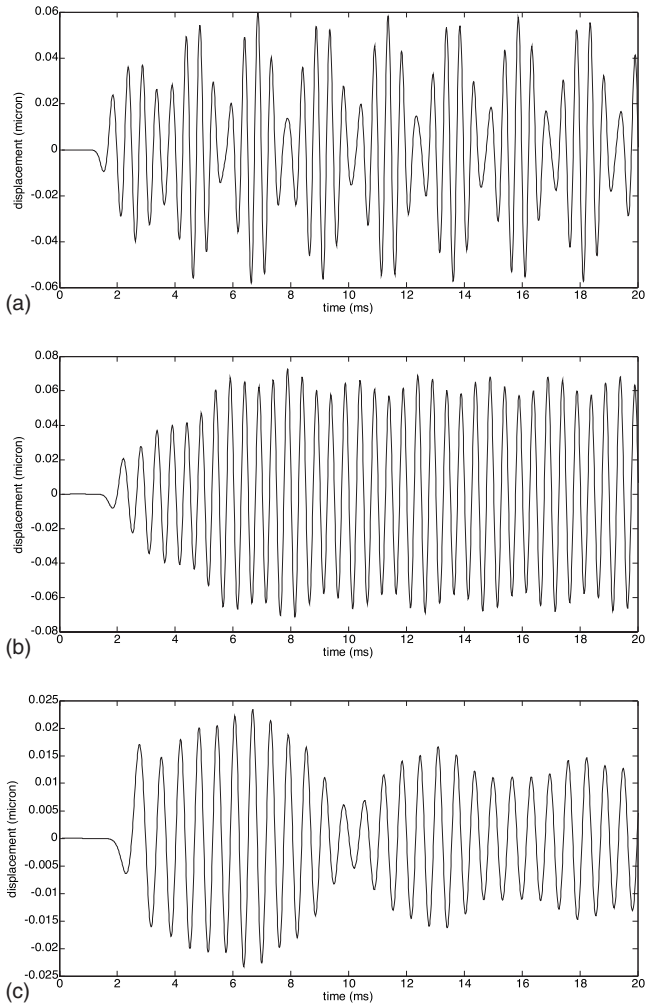


FIG. 5. Time evolution of the cochlear response at the three cochlear places of Figs. 4(a)–4(c).

time-domain solution permits a direct estimate of the response waveform at the base (and at all other cochlear places) allowing us to compute time delays directly, without any assumption about the linearity of the system.

B. DPOAEs

In Fig. 4, we show the generation of the $2f_1-f_2$ distortion product due to nonlinear interaction of two primary tones ($f_1=2000$ Hz, $f_2/f_1=1.22$, $L_1-L_2=10$ dB, and $L_2=60$ dB SPL). In Fig. 4, the spectrum of the cochlear displacement is shown at different cochlear positions. The two primary tones propagate up to $x(f_2)$, where the DPOAE is generated and the f_2 tone is absorbed (and partially reflected by roughness) [Fig. 4(a)], then the f_1 tone is absorbed (and partially reflected by roughness) at its resonant place [Fig. 4(b)], whereas the distortion tone propagates forward to its tonotopic place, where it is amplified [Fig. 4(c)], absorbed, and partially reflected by roughness. The continuum spectrum, shifting to lower frequencies with increasing x, which can be observed below the spectral lines, is due to a small spurious broad-band TW. Several distortion product lines are visible in the spectrum of the response at the stapes [Fig. 4(d)], the most intense being that of frequency $f_{DP}=2f_1-f_2$, which is about 30 dB below the primary stimulus level. These DP levels are rather high, which is an indication that the parameters of the model still need to be optimized. At the present stage, a high level of DPOAE response may help show the qualitative behavior of the model.

The time-domain solution allows one to follow the generation of the DPOAE response also looking at displacement at fixed cochlear positions x as a function of time or at fixed times as a function of the position x. At the same three cochlear places of Figs. 4(a)–4(c), one gets the time evolution shown in Figs. 5(a)–5(c). From these plots, one can visually appreciate the different onset times of the response at different cochlear positions and the different frequency contents of the signal.

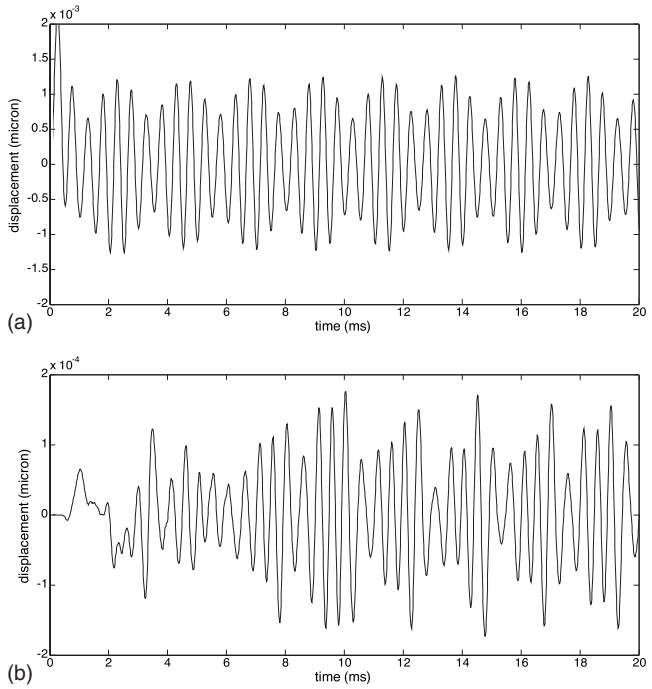


FIG. 6. Response at the stapes of the simulation shown in Fig. 4. which is obviously dominated by the stationary intense primary tones (a). Performing an identical simulation in which the nonlinear distortion term was suppressed in the region around $x(f_2)$ and subtracting the two response waveforms at the stapes, one can directly appreciate the onset delay of the DP at the base in the time domain (b).

In Fig. 6(a), we show the time-domain response at the stapes, which is obviously dominated by the intense primary tones. Performing an identical simulation in which the nonlinear distortion term is suppressed in the region around $x(f_2)$ and subtracting the two response waveforms, one can cancel the contribution from the stimuli and appreciate the onset delay of the DP at the base in the time domain [Fig. 6(b), note that the scale is ten times smaller than that in Fig. 6(a)]. The observed onset delay is compatible with the forward transmission delay of the primaries from the base to $x(f_2)$ estimated from the TEOAE simulation, plus a shorter backward transmission delay of the DP, as predicted by theory, and similar to what has been actually observed by experimental studies of the DPOAE onset time (Whitehead *et al.*, 1996). The backward delay is expected to be shorter, as explained in Sec. II, because the frequency of the DP is lower than the characteristic frequencies of the backward cochlear path; therefore, its propagation is faster that that of the f_2 tone along the same forward path. Slightly later, a contribution of the primary tones from their tonotopic places is expected to reach the base. This contribution is not canceled by the subtraction technique because, having suppressed nonlinear damping in the second simulation, the f_1 and f_2 components do not cancel exactly. The DPOAE contribution from the second source would come back even later, due to its lower frequency and level.

V. CONCLUSIONS

The multi-iterative computational strategies used within a stable time-step integrator based on implicit formulas considered in this study allowed us to solve accurately and efficiently our full cochlear model in the time domain. This is important in order to study the characteristic time delays associated with the propagation of acoustic signals along the BM, removing the ambiguities associated with the use of frequency-domain formulations, which are fully meaningful only for linear systems. A new 1D model, including feedforward nonlinear and non-local terms, as well as cochlear roughness and a middle ear equation, has been implemented in a matrix formulation scheme, proposed by Elliott *et al.* (2007).

The results show that several aspects of the OAE phenomenology can be effectively predicted by such a model formulation, and help to design specific experiments dedicated to the study of a specific issue. In particular, the TEOAE latency/frequency relation is predicted in fine agreement with experimental data, and the DPOAE onset latency is shown to be associated with the BM forward and backward transmission delays, with results comparable to those of experimental studies on the DPOAE onset time.

Some parameters of the proposed continuous full cochlear model still need to be refined and tuned with an accurate comparison of its prediction with all available experimental data. After that, this model formulation can be extensively used to design experimental campaigns and diagnostic techniques, and to interpret the results of new experiments.

Abdala, C., and Folsom, R. C. (**1995**). "Frequency contribution to the click-evoked auditory brain-stem response in human adults and infants," J. Acoust. Soc. Am. **97**, 2394–2404.
Bertaccini, D., and Fanelli, S. (**2009**). "Computational and conditioning issues of a discrete model for sensorineaural hypoacusia," Appl. Numer. Math. **59**, 1989–2001.
Choi, Y., Lee, S., Parham, K., Neely, S. T., and Kim, D. O. (**2008**). "Stimulus-frequency otoacoustic emissions: Measurements and simulations with an active cochlear model," J. Acoust. Soc. Am. **123**, 2651–2669.
de Boer, E., and Nuttall, A. L. (**2003**). "Properties of amplifying elements in the cochlea," in *Biophysics of the Cochlea: From Molecules to Models*, edited by A. W. Gummer (World Scientific, Singapore), pp. 331–342.
de Boer, E., Zheng, J., Porsov, E., and Nuttall, A. L. (**2008**). "Inverted direction of wave propagation (IDWP) in the cochlea," J. Acoust. Soc. Am. **123**, 1513–1521.
Don, M., Ponton, C. W., Eggermont, J. J., and Masuda, A. (**1993**). "Gender differences in cochlear response time: An explanation for gender amplitude differences in the unmasked auditory brainstem response," J. Acoust. Soc. Am. **94**, 2135–2148.
Donaldson, G. S., and Ruth, R. A. (**1993**). "Derived band auditory brainstem response estimates of traveling wave velocity in humans. I: Normal-hearing subjects," J. Acoust. Soc. Am. **93**, 940–951.
Dong, W., and Olson, E. S. (**2008**). "Evidence for reverse cochlear traveling waves," J. Acoust. Soc. Am. **123**, 222–240.
Eggermont, J. J., and Don, M. (**1980**). "Analysis of the click-evoked brainstem potentials in humans using high-pass noise masking. II. Effect of click intensity," J. Acoust. Soc. Am. **68**, 1671–1675.
Elliott, S. J., Ku, E. M., and Lineton, B. (**2007**). "A state space model for cochlear mechanics," J. Acoust. Soc. Am. **122**, 2759–2771.
Furst, M., and Lapid, M. (**1988**). "A cochlear model for acoustic emissions," J. Acoust. Soc. Am. **84**, 222–229.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Greenwood, D. D. (**1990**). "A cochlear frequency position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.
He, W., Nuttall, A. L., and Ren, T. (**2007**). "Two-tone distortion at different longitudinal locations on the basilar membrane," Hear. Res. **228**, 112–122.
Jedrzejczak, W. W., Blinowska, K. J., Konopka, W., Grzanka, A., and Durka, P. J. (**2004**). "Identification of otoacoustic emission components by

means of adaptive approximations," J. Acoust. Soc. Am. **115**, 2148–2158.
Kim, J., and Xin, J. (**2005**). ""A two-dimensional nonlinear nonlocal feedforward cochlear model and time domain computation of multitone interactions," Multiscale Model. Simul. **4**, 664–690.
Lim, K. M., and Steele, C. R. (**2002**). "A three-dimensional nonlinear active cochlear model analyzed by the WKB-numeric method," Hear. Res. **170**, 190–205.
Long, G. R., Talmadge, C. L., and Lee, J. (**2008**). "Measuring distortion product otoacoustic emissions using continuously sweeping primaries," J. Acoust. Soc. Am. **124**, 1613–1626.
Moleti, A., and Sisto, R. (**2003**). "Objective estimates of cochlear tuning by otoacoustic emission analysis," J. Acoust. Soc. Am. **113**, 423–429.
Neely, S. T., and Kim, D. O. (**1986**). "A model for active elements in cochlear biomechanics," J. Acoust. Soc. Am. **79**, 1472–1480.
Neely, S. T., Norton, S. J., Gorga, M. P., and Jesteadt, W. (**1988**). "Latency of auditory brain-stem responses and otoacoustic emissions using tone-burst stimuli," J. Acoust. Soc. Am. **83**, 652–656.
Nobili, R., and Mammano, F. (**1996**). "Biophysics of the cochlea II: Stationary nonlinear phenomenology," J. Acoust. Soc. Am. **99**, 2244–2255.
Prijs, V. F., Schneider, S., and Schoonhoven, R. (**2000**). "Group delays of distortion product otoacoustic emissions: Relating delays measured with f1- and f2-sweep paradigms," J. Acoust. Soc. Am. **107**, 3298–3307.
Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2027–2067.
Puria, S. (**2003**). "Measurements of human middle ear forward and reverse acoustics: Implications for otoacoustic emissions," J. Acoust. Soc. Am. **113**, 2773–2789.
Ren, T. (**2004**). "Reverse propagation of sound in the gerbil coclea," Nat. Neurosci. **7**, 333–334.
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.
Schoonhoven, R., Prijs, V. F., and Schneider, S. (**2001**). "DPOAE group delays versus electrophysiological measures of cochlear delay in normal human ears," J. Acoust. Soc. Am. **109**, 1503–1512.
Shera, C. A., and Guinan, J. J., Jr. (**1999**). "Evoked otoacoustic emissions arise from two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.
Shera, C. A., and Guinan, J. J., Jr. (**2003**). "Stimulus-frequency emission group delay: A test of coherent reflection filtering and a window on cochlear tuning," J. Acoust. Soc. Am. **113**, 2762–2772.
Shera, C. A., Guinan, J. J., Jr., and Oxenham, A. J. (**2002**). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," Proc. Natl. Acad. Sci. **99**, 3318–3323.
Shera, C. A., Tubis, A., and Talmadge, C. L. (**2005**). "Coherent reflection in a two-dimensional cochlea: Short-wave versus long-wave scattering in the generation of reflection-source otoacoustic emissions," J. Acoust. Soc. Am. **118**, 287–313.
Shera, C. A., Tubis, A., and Talmadge, C. L. (**2006**) "Delays of SFOAEs and cochlear vibrations support the theory of coherent reflection filtering," Association for Research in Otolaryngology 2006 Meeting Poster.
Shera, C. A., Tubis, A., Talmadge, C. L., de Boer, E., Fahey, P. F., and Guinan, J. J., Jr. (**2007**). "Allen–Fahey and related experiments support the predominance of cochlear slow-wave otoacoustic emissions," J. Acoust. Soc. Am. **121**, 1564–1575.
Shera, C. A., and Zweig, G. (**1991**). "Reflection of retrograde waves within the cochlea and at the stapes," J. Acoust. Soc. Am. **89**, 1290–1305.
Siegel, J. H., Cerka, A. J., Recio-Spinoso, A., Temchin, A. N., van Dijk, P., and Ruggero, M. A. (**2005**). "Delays of stimulus-frequency otoacoustic emissions and cochlear vibrations contradict the theory of coherent reflection filtering," J. Acoust. Soc. Am. **118**, 2434–2443.
Sisto, R., and Moleti, A. (**2002**). "On the frequency dependence of the otoacoustic emission latency in hypoacoustic and normal ears," J. Acoust. Soc. Am. **111**, 297–308.
Sisto, R., and Moleti, A. (**2007**). "Transient evoked otoacoustic emission latency and cochlear tuning at different stimulus levels," J. Acoust. Soc. Am. **122**, 2183–2190.
Sisto, R., Moleti, A., and Shera, C. A. (**2007**). "Cochlear reflectivity in transmission-line models and otoacoustic emission characteristic time delays," J. Acoust. Soc. Am. **122**, 3554–3561.
Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). "Modeling otoacoustic emission and hearing threshold fine structures," J. Acoust. Soc. Am. **104**, 1517–1543.
Talmadge, C. L., Tubis, A., Long, G. R., and Tong, C. (**2000**). "Modeling the combined effect of basilar membrane nonlinearity and roughness on stimulus frequency otoacoustic emission fine structure," J. Acoust. Soc. Am. **108**, 2911–2932.
Tognola, G., Ravazzani, P., and Grandori, F. (**1997**). "Time-frequency distributions of click-evoked otoacoustic emissions," Hear. Res. **106**, 112–122.
Unoki, M., Miyauchi, R., and Tan, C.-T. (**2007**). "Estimates of tuning of auditory filter using simultaneous and forward notched-noise masking," in *Hearing—From Sensory Processing to Perception*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Springer, Berlin), pp. 19–26.
Voss, S. E., and Shera, C. A. (**2004**). "Simultaneous measurement of middle-ear input impedance and forward/reverse transmission in cat," J. Acoust. Soc. Am. **116**, 2187–2198.
Whitehead, M. L., Stagner, B. B., Martin, G. K., and Lonsbury-Martin, B. L. (**1996**). "Visualization of the onset of distortion-product otoacoustic emissions, and measurement of their latency," J. Acoust. Soc. Am. **100**, 1663–1679.
Zweig, G., and Shera, C. A. (**1995**). "The origin of periodicity in the spectrum of otoacoustic emissions," J. Acoust. Soc. Am. **98**, 2018–2047.

Selective filtering to spurious localization cues in the mammalian auditory brainstem

Hamish Meffin[a] and Benedikt Grothe
Department of Biology II Neurobiology and Bernstein Center for Computational Neuroscience, Ludwig-Maximilians-University Munich, Großhaderner Strae 2, D-82152 Planegg-Martinsried, Germany

(Received 12 May 2008; revised 7 August 2009; accepted 31 August 2009)

The cues used by mammals to localize sound can become corrupted when multiple sound sources are present due to the interference of sound waves. Under such circumstances these localization cues become spurious and often fluctuate rapidly (>100 Hz). By contrast, rapid fluctuations in sound pressure level do not indicate a corrupted signal, but rather may convey important information about the sound source. It is proposed that filtering in the auditory brainstem acts to selectively attenuate signals associated with the presence of rapidly fluctuating (spurious) localization cues, but not those associated with slowly varying cues. Further it is proposed that specific inhibitory circuitry in the auditory brainstem, centered on the dorsal nucleus of the lateral lemniscus (DNLL), contributes to this selective filtering. Data from extra-cellular recordings in anesthetized Mongolian gerbils are presented to support these hypotheses for a subpopulation of DNLL neurons. These results provide new insights into how the mammalian auditory system processes information about multiple sound sources. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3238239]

PACS number(s): 43.64.Qh [RYL] Pages: 2437–2454

I. INTRODUCTION

In mammals, the localization of a single, low-frequency (<3 kHz) sound in the horizontal plane is based primarily on the difference in time at which the sound arrives at each ear (Rayleigh, 1907; Erulkar, 1972; Macpherson and Middlebrooks, 2002). At these frequencies the auditory system utilizes this cue, termed the interaural time difference (ITD), with great acuity by relying on the phased-locked response of auditory nerve fibers to intense spectral frequencies in the sound (Klumpp and Eady, 1956; Rose *et al.*, 1967). The critical step in extracting the ITD is achieved in the superior olivary complex via a coincidence detection mechanism that compares the phased-locked responses arriving from each ear (Jeffress, 1948; Yin and Chan, 1990; Spitzer and Semple, 1995; Brand *et al.*, 2002). A second major binaural cue available to the auditory system for localization, namely, the interaural intensity difference (IID), results from the shadowing effect of the head and pinnae (Rayleigh, 1907; Erulkar, 1972; Macpherson and Middlebrooks, 2002). This cue is important at high frequencies (>1.5–5 kHz, depending on head size), for which the shadowing effect is great. For low frequencies, however, IIDs are minimal except for near field sounds (Shinn-Cunningham *et al.*, 2000).

When sounds arrive simultaneously from spatially separated sources, as is common in noisy or reverberant environments, binaural localization cues may become corrupted due to the interference of the different sound waves (Bauer, 1961; Takahashi and Keller, 1994; Blauert, 1997; Roman *et al.*, 2003; Keller and Takahashi, 2005; see Fig. 1 and discussion below). We refer to such localization cues as spurious localization cues, or spurious cues for short. A characteristic of these spurious cues, which is common for many natural sounds, is that they fluctuate rapidly: on the same time scale as the amplitude and phase modulations in the sound sources (e.g., >100 Hz; see below and Fig. 1). This phenomenon is known as interaural decorrelation since the left and right ear sound waves are imperfectly correlated in this situation. In this circumstance, the sensitivity of human listeners to the *localization* information normally inherent in these binaural cues is reduced (Blauert, 1972; Grantham and Wightman, 1978). However these spurious cues still contribute to the overall perception of the sound and may even assist with the analysis of the auditory scene in other ways under some circumstances, e.g., the *detection* of a spatial separated sound source in a noisy environment, in which the signal to noise ratio is poor (Hirsh, 1948; Durlach and Colburn, 1978).

The rapid fluctuations typical of spurious ITDs and IIDs are suggestive of a solution for how the auditory system may deal with these corrupted cues, but they also give rise to a difficulty. On the one hand, such fluctuations in spurious ITDs and IIDs are far more rapid than those caused by the physical movement of either a typical source or receiver. This difference between fluctuation rates leads to the possibility that the auditory system may deal with the spurious cues by simply suppressing the rapid fluctuations with a low-pass filter. This could, for example, lead to a degree of suppression of the spike-rate of ITD-sensitive neurons and/or to a degree of suppression of ITD tuning. This is consistent with the observation that human listeners are unable to follow the location of a sound when ITD and IID fluctuate at even modest rates (as small as 2–3 Hz, and certainly for rates of 20 Hz and above) (Blauert, 1972; Grantham and Wightman, 1978). Further, it could, in principle, still permit these cues to contribute to the overall perception of a sound

[a]Author to whom correspondence should be addressed. Present address: NICTA VRL, c/-Department of Electrical and Electronic Engineering, University of Melbourne, Victoria 3010, Australia. Electronic mail: hmeffin@yahoo.com

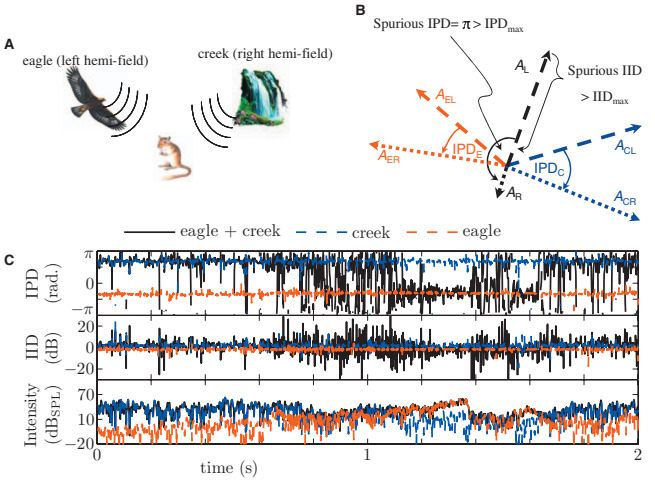


FIG. 1. Illustration of the interference of sound waves leading to the production of spurious IPDs and IIDs. (A) Sounds arrive at left and right ears (of a gerbil) from spatially separated sources: here an eagle (E) and a creek (C). (B) The interference of sound waves from two sources leading to spurious IPDs and IIDs is illustrated for a single frequency channel using a vector representation. A contribution from a given source (s=E or C) reaching a particular ear (e=L or R) is represented graphically by a two-dimensional vector, whose magnitude, A_{se}, and phase match those of the sinusoid for that frequency channel. Each of the two sources (eagle in red and creek in blue) is represented by a left (dashed) and right (dotted) vector. The IPDs for each source alone (IPD_E or IPD_C) are marked. The vectors for the composite waveforms are shown in black, as either dashed (left) or dotted (right) lines. The resulting spurious IPD and IID are marked. (C) Separate recordings of two sounds (an eagle and a creek) were taken, and ITDs and IIDs were introduced from the Mongolian gerbil's head related transfer function so that the sources appeared to come from opposite hemi-fields. The plots show the temporal fluctuations in the 1.5-kHz frequency band for IPD, IID, and intensity for the eagle alone (red), the creek alone (blue), and the eagle+steam composite (black).

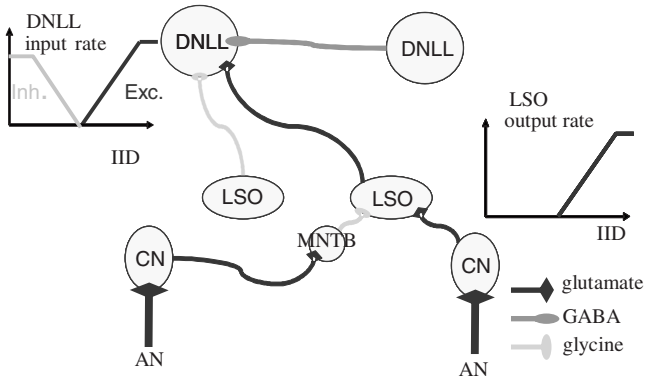


FIG. 2. Schematic showing the circuitry underlying the complementary excitatory and persistently inhibitory IID tuning of a DNLL IID-sensitive neuron. For clarity, mirror symmetric projections are not shown. The complementary excitatory and inhibitory input rate to the left DNLL is shown in the plot, adjacent to the left DNLL, as a function of IID. Theoretically, the circuit is capable of selectively filtering out rapidly fluctuating IIDs while letting rapidly fluctuating changes in binaurally coherent amplitude pass. Abbreviations: AN, auditory nerve; CN, cochlear nucleus; MNTB, medial nucleus of the trapezoid body; LSO, lateral superior olive; and DNLL, dorsal nucleus of the lateral lemniscus.

since the neural code has been altered relative to the response it would produce from non-fluctuating spatial cues. On the other hand, a difficulty with this solution is that most ITD/IID-sensitive neurons in the auditory brainstem also follow the rapid (binaurally coherent) amplitude modulations in a sound (Joris *et al.*, 2004). Consequently, applying a linear low-pass filter to such co-modulated input would necessarily suppress not only the spurious ITDs and IIDs but also the amplitude modulations. This may be deleterious during periods when a single source dominates and the amplitude modulations may provide behaviorally important information. In this case, it may be useful to have a filter that selectively suppresses response when there are rapidly fluctuating ITDs and IIDs, but not when there are just rapid modulations in binaurally coherent amplitude/intensity (BCI=instantaneous intensity averaged across the two ears). We will refer to this idea as the selective filtering hypothesis.

Keller and Takahashi (2005) examined the issue of selective filtering in the barn owl. They advanced the hypothesis that neurons integrate information about ITDs and IIDs across all or parts of the tonotopic frequency band. This allows a neuron to respond strongly when the tonotopic spread of ITDs and IIDs corresponds to a single spatial position in its receptive field. However, during periods when spurious binaural localization cues prevail, it results in a weak response because IIDs and ITDs have incoherent values across the tonotopic array. Further, there is no problem with losing sensitivity to binaurally coherent amplitude modulations in this model because integration occurs across tonotopic frequency rather than across time, and amplitude modulations are assumed to be largely consistent across tonotopic frequency. A difficulty with this scheme may occur if there are multiple sound sources with non-overlapping frequency content. Then integration across frequency may give a weak response, despite valid cues, due to the different sources having ITDs and IIDs that are not coherent with each other. This problem might be avoided by integrating across multiple different subsets of the frequency spectrum in different neurons. Other studies have addressed the related issue of echo suppression and particularly the suppression of information about the location of an echo, but have not considered the issue of selective filtering by spurious localization cues that is examined here (Fitzpatrick *et al.*, 1995; Fitzpatrick *et al.*, 1999; Litovsky and Yin, 1998a, 1998b; Burger and Pollak, 2001; Pecka *et al.*, 2007).

In this study, we hypothesize that a neural circuit, centered on the dorsal nucleus of the lateral lemniscus (DNLL; Fig. 2) in the auditory brainstem, contributes to a selective filtering function via a so-called "persistent suppression" (Yang and Pollack, 1994; Pollack, 2003; Pecka *et al.*, 2007). In the remaining part of the Introduction, further background and explanation for this hypothesis are presented.

The production of spurious ITDs and IIDs occurs when simultaneous sounds arrive from spatially separated sources and have similar spectral energy within a given frequency channel as, for example, the call of an eagle (E) and the burbling of a creek (C) (Fig. 1). It is most pronounced when the sources are in opposite hemi-fields. The underlying mechanism is the interference of sound waves, which can be understood in terms of the vector sum of single channel Fourier components [Fig. 1(b)]. For any given frequency channel (f) the contribution from a given source (s=E or C) reaching a particular ear (e=L or R) is given by $A_{s,e}\cos(2\pi ft+\phi_{s,e})$, where the amplitude $A_{s,e}$ and phase $\phi_{s,e}$ are generally func-

tions of time that vary at frequencies that are much less than the channel frequency. Such a contribution may be represented graphically by a two-dimensional vector, $\boldsymbol{v}_{s,e}$, of length $A_{s,e}$ and phase $\phi_{s,e}$.

In Fig. 1(b) each of the two sources (eagle in red and creek in blue) is represented by vectors for the left (dashed) and right (dotted) ears. It is convenient here, and throughout, to normalize the ITD for the period of the channel frequency, in which case one obtains the interaural phase difference: $\mathrm{IPD}=2\pi f\times\mathrm{ITD}$. The IPD for each source alone (IPD_E or IPD_C) is represented by the angle between the relevant pair of left and right vectors. The IID is related to the ratio of vector lengths [$\mathrm{IID}_s=20\log_{10}(A_{s,\mathrm{R}}/A_{s,\mathrm{L}})$]. Notice that, in this example, IPD_E and IPD_C have opposite signs since the two sound sources are in opposite hemi-fields [i.e., the arrows indicating the angles in Fig. 1(b) point in opposite directions]. Also both IIDs are negligible, which is typical at the low frequencies for which IPDs are used for localization. In general, the phase angle between two independent sources is arbitrary and randomly varying over time. The composite waveforms are given by the vector sum of the two sources at each ear [$\boldsymbol{v}_e=\boldsymbol{v}_{E,e}+\boldsymbol{v}_{C,e}$; black vectors that are either dashed (left) or dotted (right) in Fig. 1(b)].

When the two sources have similar amplitudes, significant cancellation can occur, resulting in composite waveforms with an IPD and IID that bear no clear relation to the IPD and IID of either source, as is the case in this example. Indeed, the composite IPD and IID are influenced by additional variables, such as the relative phase and amplitudes of sound waves between sources (rather than between ears). Consequently, the composite IPD and IID are corrupted as localization cues since information about the additional variables is required to accurately infer location. Thus, the cues are termed spurious. Furthermore, the spurious IPD and IID may be (and in this example are) greater than their respective single sound source limits as imposed by head size for IPD and the maximal degree of head shadowing for IID (Maki and Furakawa, 2005).

Figure 1(c) shows the rapid (>100 Hz) fluctuations in spurious IPDs and IIDs during a period in which both the eagle call and the creek burbling had similar spectral energy within the 1.5-kHz frequency channel (1.44–1.56 kHz) between 0.7 and 1.1 s after the sounds commence. These fluctuations were caused by the underlying intensity and phase fluctuations in the two sounds [Fig. 1(c), bottom plot], which contained significant spectral energy up to frequencies of more than a hundred hertz. Such a frequency range for amplitude modulation is common in sounds. In comparison, large fluctuations in IPD and IID caused by the physical movement of either source or receiver are typically limited to have spectral energy below 10 Hz. (For instance, even in the highly unlikely event of an object passing tangentially, directly in front of a listener at 100 km/h at a distance of 1 m, it would take 0.14 s to move from 45° right to 45° left and back again, producing IPD and IID fluctuations up to only 7 Hz).

The hypothesized role of persistent inhibition in selective filtering by spurious IPDs and IIDs can be understood by considering the neural circuit shown in Fig. 2. Persistent inhibition has been reported previously in DNLL neurons sensitive to IID in the context of echo suppression (Yang and Pollak, 1994a, 1994b; Burger and Pollak, 2001; Pecka *et al.*, 2007). These DNLL neurons inherit their characteristic sigmoidal IID tuning via an excitatory projection from the contralateral lateral superior olive (LSO) (Pecka *et al.*, 2007) (Fig. 2: plot adjacent to the right LSO). The persistent inhibition arises predominantly from a GABA-ergic projection from the contralateral DNLL, whose tuning is essentially complementary to the excitatory tuning (Burger and Pollak, 2001; Pecka *et al.*, 2007) (Fig. 2: also, note the additional glycinergic input from the ipsilateral LSO). Consequently, and somewhat surprisingly, the inhibitory input to the DNLL neuron acts only at those IIDs for which excitation is absent. The defining feature of the "persistent" GABA-ergic inhibition is its long time course compared to that of the other inputs, such that a sound with an inhibitory IID can suppress any induced neural response for an average of 20 ms following the sound's cessation (Yang and Pollak, 1994a, 1994b).

We hypothesize that this so-called persistent inhibition may act to selectively filter the spike-rate during rapid (>100-Hz) IID fluctuations because such fluctuations will move rapidly between the inhibitory and excitatory subfields of the neuron's IID tuning curve. Every time the IID enters the inhibitory subfield, it will invoke the inhibition, which will persist as it re-enters the excitatory subfield, leading to a relative suppression of neural spike-rate. In contrast, during slow IID fluctuations (<10 Hz), the persistent inhibition will have finished by the time the IID re-enters the excitatory subfield, so there will be no suppression. Likewise, rapid fluctuations in BCI will not invoke the persistent inhibition, provided that they occur in combination with an excitatory IID. Consequently the neuron will be able to follow the modulations unsuppressed via the fast excitatory synapse. Notice that the mechanism will produce the greatest effect when sources are in opposite hemi-fields, which is also when IPDs and IIDs become most heavily corrupted. Note also that a similar mechanism could work for rapidly fluctuating, spurious IPDs.

II. MATERIALS AND METHODS

In this study, we preformed experiments with two sets of stimuli for each animal. The first set was designed to test the selective filtering hypothesis, while the second set was used to measure persistent suppression. The presentation of these two stimuli sets will be referred to as Experiment 1 and Experiment 2, respectively, throughout the rest of the paper.

A. Acoustic stimuli

1. Experiment 1: Selective filtering

Stimuli used to directly test the selective filtering hypothesis were narrowband noise centered on the neuron's best frequency (BF; frequency of maximum response 30 dB above threshold). In each stimulus, the binaural localization cues (either IPD or IID) were manipulated to randomly fluctuate either "rapidly," mimicking the spurious cues, or "slowly," mimicking valid cues. The elicited neural response was then compared between rapid and slow stimuli. Here

TABLE I. The six narrowband noise stimuli used to test the selective filtering hypothesis. For each stimulus, the three binaural cues of IPD, IID, and BCI independently had either rapid (r) or slow (s) fluctuations, as indicated in the table. IPS was slow for all stimuli.

Notation	IPD	IID	BCI
S(IPD=s,IID=s,BCI=s)	Slow	Slow	Slow
S(IPD=s,IID=s,BCI=r)	Slow	Slow	Rapid
S(IPD=s,IID=r,BCI=s)	Slow	Rapid	Slow
S(IPD=r,IID=s,BCI=s)	Rapid	Slow	Slow
S(IPD=s,IID=r,BCI=r)	Slow	Rapid	Rapid
S(IPD=r,IID=s,BCI=r)	Rapid	Slow	Rapid

slowly means most fluctuations occurred at rates $\leqslant$20 Hz, while rapidly means that most fluctuations occurred at rates $\leqslant r$, with r between 166 and 1000 Hz, depending on the neuron's BF (further details are given below). In addition, to test the selectivity of the filtering, control stimuli were also presented for which the BCI randomly fluctuated either rapidly or slowly at rates that were the same as the corresponding IPD/IID rates. According to the selective filtering hypothesis, we expected to see an attenuation of response for rapid compared to slow stimuli only in the case of IPD and IID, but not for BCI. Below, the rationale for the choice of stimuli is first explained, followed by a description of their synthesis.

Six stimuli were used, as listed in Table I. These differed in the particular combination of cues that were chosen to vary rapidly or slowly. For example, the stimulus labeled S(IPD=s,IID=s,BCI=r) had slowly varying ("s") IPD and IID, but rapidly varying ("r") BCI. Figure 3(a) shows an example of the cue waveforms for this particular type of stimulus. Importantly, the stimuli differed little in their overall distributions of IPD, IID, and BCI, as calculated for the whole stimulus. Consequently, any significant differences in the response to the stimuli were due to differences in rates of fluctuation of one or more of the binaural cues. The distributions were as follows: IPD was approximately uniformly distributed between $-\pi$ and π, IID was approximately Gaussian distributed with mean 0 dB and standard deviation 14 dB, and BCI was approximately Gaussian distributed with mean 60–70-dB sound pressure level (SPL), depending on neural threshold, and standard deviation 14 dB.

The use of the six stimuli allowed for three independent tests of the selective filtering hypothesis, as detailed in Table II. The tests in each of the three rows involved the comparison of neural response for different pairs of probe and reference stimuli. Each pair allowed the comparison of rapid vs slow fluctuations in either IPD, IID, or BCI, depending on which pair of Probe and Reference stimuli were chosen. The Reference stimuli were the same across all pairs within a test set (i.e., within a row in Table II). For example, the test set in the first row was labeled "s Ref (all)" because the Reference stimulus, S(IPD=s,IID=s,BCI=s), had slow fluctuations for all binaural cues. To compare rapid vs slow fluctuations in IPD, the corresponding probe stimulus was S(IPD=r,IID=s,BCI=s) because this stimulus had rapidly fluctuating IPD, but was otherwise the same as the Reference. Similar logic applies for the probe stimuli listed for IID and BCI. Note also that the three cues were treated symmetrically with respect to the relationship between probe and reference stimuli (i.e., the relationship always corresponded to rapid vs slow in this test set). In principle, this was important in order to test the selective filtering hypothesis in an unbiased fashion when assessing the relative effect of rapid vs slow fluctuations in IPD, IID, and BCI.

The test sets in the second and third rows were similar, but with some differences worth noting. The test set, "r Ref (IPD, BCI)," in row 2 was designed as an additional test for the selective filtering with respect to IPD. It had the stimulus S(IPD=r,IID=s,BCI=r) as reference with rapidly fluctuating IPD and BCI. Given the requirement that Probe and Reference should have a symmetric relationship across the test set (i.e., r vs s, or s vs r), this allowed only two Probes, S(IPD=s,IID=s,BCI=r) and S(IPD=r,IID=s,BCI=s), for this set. This still allowed us to compare responses for rapid vs slow fluctuations of IPD and BCI and assess whether the response to IPD showed suppression relative to that for BCI. The test set, "r Ref (IID, BCI)," in row 3 was similar to that in row 2, but it was designed to test for selective filtering with respect to IID.

The principle by which binaural narrowband noise stimuli were constructed is as follows. Given a set of time varying binaural cues consisting of BCI (=half the interaural intensity sum), IID, interaural phase sum (IPS), and IPD, a pair of left and right narrowband sound waves, $(s_L(t), s_R(t))$, was constructed to provide this set of cues

$$s_{R/L}(t) = A_{R/L}(t)\cos(2\pi ft + \phi_{R/L}(t)), \tag{1}$$

where f is the carrier frequency (=BF of the neuron defined at the start of this section), t is time, $A_{R/L}(t)$ are the right and left amplitudes, and $\phi_{R/L}(t)$ are the right and left phases. This was accomplished by choosing the amplitudes and phases to be functions of BCI, IID, IPS, and IPD as follows:

$$A_R(t) = 10^{\mathrm{BCI}(t)/20+\mathrm{IID}(t)/40}, \tag{2}$$

$$A_L(t) = 10^{\mathrm{BCI}(t)/20-\mathrm{IID}(t)/40}, \tag{3}$$

$$\phi_R(t) = \frac{\mathrm{IPS}(t)}{2} + \frac{\mathrm{IPD}(t)}{2}, \tag{4}$$

$$\phi_L(t) = \frac{\mathrm{IPS}(t)}{2} - \frac{\mathrm{IPD}(t)}{2}. \tag{5}$$

Equations (2)–(5) can be found by inverting the following standard formulas for BCI, IID, IPS, and IPD to obtain A_R, A_L, ϕ_R, and ϕ_L:

$$\mathrm{BCI} = 10\log_{10}(A_R A_L), \tag{6}$$

$$\mathrm{IID} = 20\log_{10}(A_R/A_L), \tag{7}$$

$$\mathrm{IPS} = \phi_R + \phi_L, \tag{8}$$

$$\mathrm{IPD} = \phi_R - \phi_L. \tag{9}$$

Alternatively Eqs. (2)–(5) can be verified as correct by substituting them into Eqs. (6)–(9).

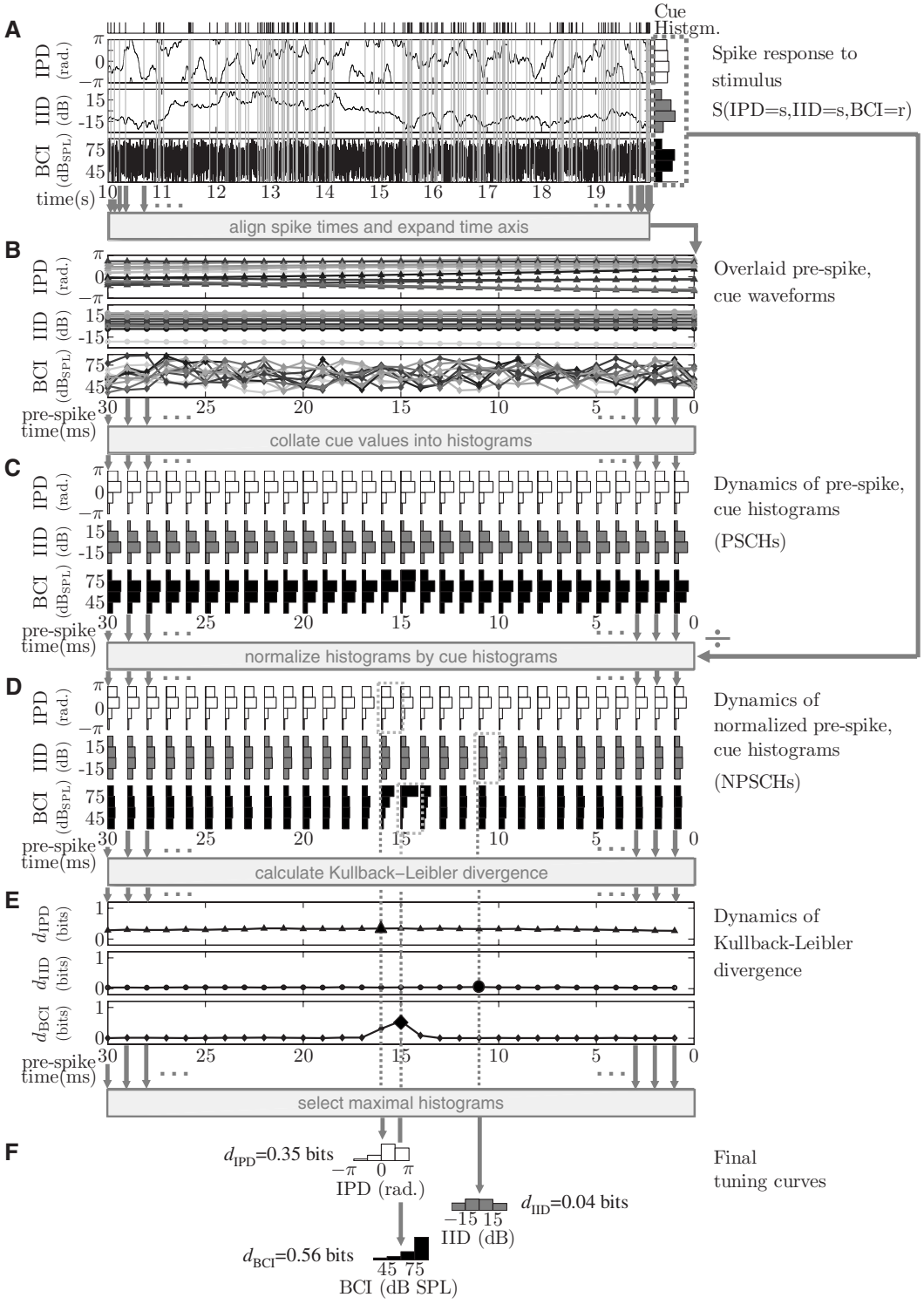


FIG. 3. Illustration of the calculation of tuning curves and tuning measures. (A) The main plots show the spike response of an X/X neuron to a section of a single sweep of the stimulus S(IPD=s,IID=s,BCI=r) (see Table I). Spikes (upper trace and gray vertical lines) are shown against the fluctuating binaural stimulus cues (black; IPD, IID, and BCI as labeled). On the right are stimulus histograms showing the total time that a given value prevailed during the stimulus for each binaural cue. (B) Overlaid pre-spike cue waveforms aligned according to spike time. Different gray shades distinguish traces but are otherwise arbitrary. (C) Dynamics of PSCHs. White, gray, and black indicate histograms for IPD, IID, and BCI, respectively. (D) Dynamics of normalized prespike cue histograms (NPSCHs). Format as per (C). (E) The temporal evolution of the Kullback–Leibler divergence, $d_{y,t}$. Enlarged symbols indicate the maximum. (F) Final tuning curves corresponding to the NPSCHs with maximal divergence, d_y, amongst those in (E).

For narrowband sound waves, $A_{R/L}(t)$ and $\phi_{R/L}(t)$ are functions of time that vary more slowly than the rate given by the carrier frequency f. Here, $A_{R/L}(t)$ are defined as dimensionless variables that have been normalized relative to the pressure at the normative threshold of hearing, 2×10^{-5} Pa. Using procedures to be described next, we obtained components (BCI, IID, IPS, and IPD) that fluctuated either rapidly or slowly in a random fashion. A chosen combination of rapid and slow components (BCI, IID, IPS, and IPD) was then substituted into Eqs. (2)–(5). Substituting

TABLE II. The three comparison sets used to test the selective filtering hypothesis, as labeled in the first column. Each row lists pairs of stimuli from Table I, labeled Probe and Reference (column 2). Each Probe-Reference pair allows the comparison of rapid compared to slow fluctuations in some binaural cue, such as IPD, IID, or BCI, as labeled in columns 3–5. Notice that for any given comparison set, the Reference stimulus is the same, regardless of the binaural cue, whereas the Probe stimulus varies. For the second and third comparison sets, some Probe stimuli were not applicable (n.a.), as described in the text.

Label		IPD	IID	BCI
s Ref. (all)	Probe	S(IPD=r,IID=s,BCI=s)	S(IPD=s,IID=r,BCI=s)	S(IPD=s,IID=s,BCI=r)
	Reference	S(IPD=s,IID=s,BCI=s)	S(IPD=s,IID=s,BCI=s)	S(IPD=s,IID=s,BCI=s)
r Ref. (IPD,BCI)	Probe	S(IPD=s,IID=s,BCI=r)	n.a.	S(IPD=r,IID=s,BCI=s)
	Reference	S(IPD=r,IID=s,BCI=r)	n.a.	S(IPD=r,IID=s,BCI=r)
r Ref. (IID,BCI)	Probe	n.a.	S(IPD=s,IID=s,BCI=r)	S(IPD=s,IID=r,BCI=s)
	Reference	n.a.	S(IPD=s,IID=r,BCI=r)	S(IPD=s,IID=r,BCI=r)

these results into Eq. (1), we obtained the sound waves $(s_L(t), s_R(t))$. The synthesis was performed digitally in MATLAB.

Rapidly fluctuating components were extracted from two independent pairs of narrowband noise: one noise pair, $N1_{R/L}(t)$, to give rapidly varying IID and IPD tokens and a second noise pair, $N2_{R/L}(t)$, to give independent, rapidly varying BCI and IPS tokens. First, for any given carrier frequency [f, see Eq. (1)], four 160-s, independent Gaussian random noise tokens were generated at one of two sampling rates used by the digital signal processor (RP2, Tucker Davis Technology): either 6.1 kHz for $f<2.6$ kHz or 12.2 kHz for $2.6<f<3.0$ kHz, in compliance with the Nyquist frequency to prevent aliasing. Second, these tokens were narrowband filtered around f with filter bandwidth ranging from 10% below to 10% above the carrier frequency to give the two narrowband noise pairs $N1_{R/L}(t)$ and $N2_{R/L}(t)$. Third, to obtain the time varying amplitudes, $A^{Nk}_{R/L}(t)$, and phases, $\phi^{Nk}_{R/L}(t)$ $(k=1,2)$, of the narrowband noise pairs, the Hilbert transforms (Hahn, 1996) of $Nk_{R/L}(t)$ were taken. This yielded a result of the form $A^{Nk}_{R/L}(t)e^{i\phi^{Nk}_{R/L}(t)}e^{2\pi ift}$ (where $i=\sqrt{-1}$), from which the carrier frequency was factored out, and the amplitude and phase extracted. Finally, rapidly fluctuating functions for BCI, IID, IPS, and IPD were then obtained by substituting the R/L-pairs of amplitudes and phases into equations of the form given in Eqs. (6)–(9). This procedure gave tokens with a corner frequency of around 1/3 of the carrier frequency. (For example, for a carrier frequency of 600 Hz, the token would have had significant spectral power up to a corner frequency of 200 Hz, beyond which power would rapidly decline; for a carrier frequency of 3000 Hz, the token would have had significant spectral energy up to a corner frequency of 1000 Hz. The fact that significant fluctuations occurred over a range of frequencies up to the corner frequency, rather than at one particular frequency, is a consequence of the random, rather than periodic, nature of the fluctuations.)

Corresponding slowly fluctuating tokens were obtained from their rapidly fluctuating counterparts as follows. First, all the sample values of the rapidly fluctuating token were arranged in ascending order. Then the slowly varying tokens were created by performing a random walk through this list, with boundary conditions that were reflecting for IID and BCI and periodic for IPD and IPS. There are three notable aspects to this algorithm. First, it resulted in fluctuations that were (pseudo-)random. Second, it resulted in tokens that drew from the same distribution of values as the rapidly fluctuating tokens. This was because the random walk sampled without bias from the list of values obtained from the rapidly fluctuating tokens (i.e., it sampled uniformly). By choosing the walk to be sufficiently long (160 s) that the distribution was well sampled, tokens were obtained that had approximately the same distribution of values as the rapidly varying tokens. As a check, the difference between "rapid" and "slow" distributions was quantified by placing all the samples from a given token into one of four bins. These bins were chosen so that they would have been equiprobable given an infinitely long token (i.e., perfect sampling of the distribution). This procedure resulted in an average of only 9% difference in sample count between tokens for any given bin. Further, no bias to any particular bin was apparent in these differences across the different tokens used. The third notable aspect of the algorithm was that by changing the step size of the random walk, different fluctuation rates could be obtained. This was a consequence of the walk occurring on an ordered list: for small steps, consecutive samples in the resulting token had similar values, while for larger steps, they had comparatively dissimilar values. We used a step size that was random and distributed uniformly between $\pm 8\sqrt{N}$, where N is the number of samples in the token (equal to the total length of the walk). Empirically, this gave tokens that fluctuated with a corner frequency of around 20 Hz. This was much slower than for the rapidly fluctuating counterpart (which had a corner frequency of between 166 and 1000 Hz depending on the BF of the neuron). (Note that the $\pm 8\sqrt{N}$ dependence in step size gave a standard deviation for the random walk of $8N/\sqrt{3} \approx 4.6N$ due to the central limit theorem. Thus on average, the walk traversed the full ordered set of samples 4.6 times during the walk.) Note that IPS was chosen to vary slowly in all six stimuli listed in Table I.

2. Experiment 2: Persistent suppression

Stimuli used to test for persistent suppression were binaural tones presented at BF. IID or IPD was modulated triphasically over time so as to give 20 ms with an excitatory

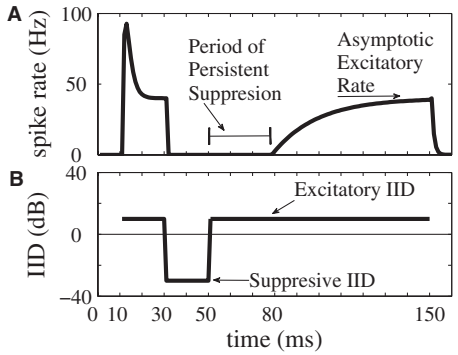


FIG. 4. Schematic illustrating the stimuli used to test for persistent suppression. (A) Characteristic spike-rate response of a persistently suppressed neuron to the stimulus. (B) Triphasic IID waveform of the stimulus. Stimulus begins at $t=10$ ms and ends at $t=150$ ms. The IID is undefined outside these times. The response in (A) has been corrected for neural propagation delay to show better alignment with the stimulus.

reference IID or IPD ($\mathrm{IID}_{\mathrm{ex}}, \mathrm{IPD}_{\mathrm{ex}}$), followed by 20 ms with a suppressive probe IID or IPD ($\mathrm{IID}_{\mathrm{sup}}, \mathrm{IPD}_{\mathrm{sup}}$), followed by another 100 ms with the original excitatory reference IID or IPD [see Fig. 4(b)]. Persistent suppression was evident as a suppression of spike-rate, relative to the asymptotic value, during the final 100-ms excitatory reference tone and following the 20-ms suppressive tone [see Fig. 4(a)]. Tests investigated either suppressive IIDs, in which case IPD=0 rad, or suppressive IPDs, in which case IID=0 dB. For suppressive IIDs, one reference and four probe IIDs were chosen such that the probes were related to the reference by $\mathrm{IID}_{\mathrm{sup}} = \{\mathrm{IID}_{\mathrm{ex}}+10\ \mathrm{dB}, \mathrm{IID}_{\mathrm{ex}}+20\ \mathrm{dB}, \mathrm{IID}_{\mathrm{ex}}+30\ \mathrm{dB}, \mathrm{IID}_{\mathrm{ex}}+40\ \mathrm{dB}\}$ or $\mathrm{IID}_{\mathrm{sup}} = \{\mathrm{IID}_{\mathrm{ex}}-10\ \mathrm{dB}, \mathrm{IID}_{\mathrm{ex}}-20\ \mathrm{dB}, \mathrm{IID}_{\mathrm{ex}}-30\ \mathrm{dB}, \mathrm{IID}_{\mathrm{ex}}-40\ \mathrm{dB}\}$. In these formulas the sign was chosen to give probe IIDs that were suppressive. The choice of sign depended on the characteristics of the neuron (for some neurons suppressive probe IIDs were found both above and below $\mathrm{IID}_{\mathrm{ex}}$, in which case they were usually investigated in two separate runs). The reference IID was chosen to be in the range −20–20 dB so as to give a non-zero firing rate, while the probe IIDs were in the range −45–45 dB and ranged well into a suppressive subfield(s) of the neuron. For suppressive IPDs, four probe IPDs were chosen to form a sequence covering $[0, 2\pi]$ separated by $\pi/2$ rad, such that one IPD evoked near-maximal suppression. The reference IPD was separated by π rad from the maximally suppressive IPD and consequently evoked near-maximal excitation. Thus, probes were related to the reference by $\mathrm{IPD}_{\mathrm{sup}} = \{\mathrm{IPD}_{\mathrm{ex}}, \mathrm{IPD}_{\mathrm{ex}} + \pi/2\ \mathrm{rad}, \mathrm{IPD}_{\mathrm{ex}} + \pi\ \mathrm{rad}, \mathrm{IPD}_{\mathrm{ex}} + 3\pi/2\ \mathrm{rad}\}$. During stimulus synthesis, modulations in IID and IPD were smoothed by a 2-ms sliding average, and the entire stimulus was bracketed by 2-ms raised-cosine ramps. Stimuli were presented between 50 and 300 times each, depending on mean spike-rate of the neuron, in a pseudo-random order.

B. Electrophysiological recordings and acoustic stimulation

The surgery used in this study has been described in detail previously (Siveke *et al.*, 2006). All experiments were approved according to the German Tierschutzgesetz (AZ 211-2531-40/01 + AZ 211-2531-68/03). Briefly, auditory responses from single neurons were recorded from 2- to 6-month-old Mongolian gerbils. Animals were anaesthetized using a physiological NaCl-solution containing ketamine (20%) and xylazine (2%) at a dosage of 0.5 ml per 100-g body weight. During surgery and while recording, a dose of 0.03 ml of the same mixture was applied subcutaneously every 20 min until the end of the experiment when animals were sacrificed without awakening by an injection of 0.1 ml of barbital (BGA-Reg. No. T331, Intervet, Germany). Constant body temperature (37–39 °C) was maintained by means of a heating pad. Skin and tissue covering the upper part of the skull were cut and pushed aside laterally. The animal was transferred to a sound-attenuated chamber where a small metal rod glued to the skull was used to secure the head of the animal in a standard stereotaxic position. For electrode penetrations to the DNLL, a small hole (<1 mm diameter) was drilled through the skull centered 1.8–1.9 mm lateral from the midline and 0.6–0.7 mm caudal of the bregma.

Single-unit responses were recorded extracellularly using glass electrodes filled with 1*M* NaCl (4–12 MΩ). Recordings were filtered and fed into a computer via an A/D converter (RP2-1, TDT). Clear isolation of responses from a single cell at a time was guaranteed by visual inspection on a spike-triggered oscilloscope and by off-line spike cluster analysis (Brainware, Jan Schnupp, TDT). For some animals, the last electrode site was marked by a pressure injection of Dextran, and the recording site was verified to be in the DNLL using standard histological techniques (for details, see Sieveke *et al.*, 2006).

Stimuli were generated by TDT System III (Tucker Davis Technologies, Alachua, FL). Digitally generated stimuli were converted to analog signals (RP2-1, TDT), attenuated (PA5, TDT), and delivered to ear-phones. The difference in the sound pressure level between the two earphones was less than 5 dB (mean 3 dB) in the range 500–3000 Hz, and the phase difference was less than 0.01 cycles. To search for acoustic responses, white noise stimuli were delivered to the contralateral ear. When a neuron was encountered, its BF and absolute threshold were determined using frequency vs level response areas. The acoustic stimuli designed to test for selective filtering and persistent suppression (described in Sec. II A) were then presented.

C. Data analysis

1. Experiment 1: Selective filtering

For the pairs of binaural noise stimuli used to test the selective filtering hypothesis (see Table II), three complementary measures were used to compare neural response when a binaural cue, x=IPD, IID, or BCI, fluctuated rapidly vs slowly. The first of these was the gain in the stimulus-averaged spike-rate for rapidly compared to slowly fluctuating stimuli:

$$g_{\mathrm{rate}}[x] = 10 \log_{10}\left(\frac{\mathrm{rate}(x=\mathrm{rapid})}{\mathrm{rate}(x=\mathrm{slow})}\right). \quad (10)$$

The other two measures were the (rate-independent) gain in tuning to a binaural cue y=IPD or BCI (tuning to IID was not calculated since IID provides little information about location at the low BFs studied here). Tuning refers here to the degree to which neural response is selective to particular values of the cue in question. In the literature, this is commonly assessed in terms of a tuning curve, which plots some measure of neural response, such as normalized spike-rate, against y. We adopted this approach as the basis for calculating the gain in tuning to a binaural cue for rapid vs slow fluctuations. The calculation was a three step process: (1) tuning curves were derived for rapid and slow stimuli, (2) the degree of tuning apparent in the tuning curves was quantified in terms of a measure (the Kullback–Leiber divergence), and (3) the gain in tuning was quantified in terms of the difference in the tuning measure between rapid and slow stimuli. This analysis is described below.

In the literature, data for tuning curves are typically obtained by presenting different stimuli, each with a different value of a cue, and measuring the response to each value. The situation in the present experiments was different because tuning curves were required from the presentation a single long (160-s) stimulus during which the cue varied constantly over a large range of values. Consequently, we used an alternative method to obtain tuning curves, in which they were constructed via a process akin to reverse correlation (Eggermont *et al.*, 1983). Traditionally, reverse correlation finds the *average* value of a stimulus cue at each time point in the period prior to the initiation of a "typical" spike. Here we used a similar process to find the full *probability distribution* for a stimulus cue at each time point in the period prior to the initiation of a typical spike. Such a distribution may be interpreted as a tuning curve since it gives the probability that a spike occurred at a particular time following the presentation of a particular cue value relative to the probability for other values of the same cue.

The process used to obtain tuning curves is illustrated in Fig. 3. Figure 3(a) shows the three fluctuating stimulus cues (IPD, IID, and BCI) as black traces relative to evoked spike times superimposed as vertical gray lines. By examining the stimulus cues in the 30-ms period prior to each spike, temporal sequences of pre-spike cue histograms (PSCHs) were constructed. This is illustrated in Figs. 3(b) and 3(c) for the three stimulus cues, y=IPD, IID, and BCI. The first step involved aligning and overlaying all 30-ms pre-spike cue fragments according to spike time [Fig. 3(b)]. Next, the prevalence of each cue value at each time slice was collated into a sequence of histograms [Fig. 3(c)]. For these histograms, time was binned into 1-ms intervals and cues had four bins with boundaries as follows: IPD=$(-\pi, -\pi/2, 0, \pi/2, \pi)$ rad; IID=$(-\infty, -15, 0, 15, \infty)$ dB; BCI =$(-\infty, 45, 60, 75, \infty)$ dB SPL. The use of just four bins allowed adequate sampling of the underlying probability distribution, yet still allowed a reasonable assessment of the tuning. Intuitively, these histograms give, for each stimulus, the frequency with which the different cue values preceded a typical spike by given time.

These histograms are not, however, proportional to the probability of a spike given a particular cue value because each cue value may occur with different probability in the stimulus (i.e., the stimuli may have a biased representation of a cue). To overcome this problem, the bins in the histograms were each divided by the corresponding *a priori* frequency of that cue value in the given stimulus. The *a priori* frequency of a cue value was calculated as the total number of seconds during the stimulus for which that particular value of the cue prevailed. Examples of these are shown as cue histograms on the right hand side of Fig. 3(a) for each of the stimulus cues, IPD, IID, and BCI. Further, it was necessary to obtain a probability function from the divided histograms by normalizing them so that they had unit sum (for purpose of calculating the Kullback–Leibler divergence, see below). These two steps yielded a sequence of normalized pre-spike cue histograms [Fig. 3(d)]. These histograms estimated the likelihood, $p_{y,t}$, that a particular value of y (=IPD, IID or BCI) preceded an arbitrary spike by time t during the stimulus in a manner that corrects for any bias in the distribution of y in the stimulus.

A tuning of neural response to y was evident in the normalized spike-triggered stimulus histograms as a favoring of some bins over others in any time slice. For example, in Fig. 3(d), the strong tuning for BCI is evident in the histograms at around 15 ms, but very little tuning is evident in the histograms at times more than a few milliseconds before or after. Mathematically, this can be expressed as a deviation of the distribution $p_{y,t}$ from the uniform distribution. On this basis, the degree of tuning to y at time t was measured using the Kullback–Leibler divergence (Kullback, 1959), which measures (here) the deviation between the neural response distribution, $p_{y,t}(j|S)(=p_{y,t})$, and the uniform distribution, $u(j)=1/N$,

$$d_{y,t}[S] = \sum_j p_{y,t}(j|S)\log_2\left(\frac{p_{y,t}(j|S)}{u(j)}\right). \tag{11}$$

Here j indexes the values of y into $N(=4)$ bins and S the stimulus for which the measure was calculated. The Kullback–Leibler divergence is derived from information theory and is related here to the degree of predictability that a spike was due to a particular cue value. It has a maximal value of 2 bits (here, due to $N=4$ bins) when $p_{y,t}$ is concentrated in a single bin (i.e., maximal tuning) and has a minimal value of 0 bit when $p_{y,t}$ is uniformly distributed (i.e., no tuning). In this context, the Kullback–Leibler divergence is similar to the synchronization index, commonly used to measure the degree of phase locking in auditory neuroscience: they both measure the degree of predictability evident in a probability distribution. The advantage of the Kullback–Leibler divergence is that it does not require a probability distribution with a periodic random variable. This allowed us to use the same measure for IPD and BCI tuning.

In Fig. 3(e) $d_{y,t}$ is plotted for each cue, y. For rapidly fluctuating cues, the response of neurons often showed a pronounced peak in $d_{y,t}$, amidst near zero values, at some time during the 30-ms pre-spike interval [e.g., at around 15 ms for y=BCI in Fig. 3(e)]. We interpreted this as the neuron responding in a tuned fashion to a brief ($\approx$3–5-ms) window of the stimulus cue, following a neural delay. For slowly

fluctuating stimuli, no pronounced peak was possible (or evident) because the cues themselves varied over a time scale longer than 30 ms. Nonetheless, a tuning was often evident as a non-zero value of $d_{y,t}$, though its value changed very little over the 30 ms. Consequently, for both fast and slow stimuli, we took the maximum in $d_{y,t}$ over the 30 ms to be a true measure of the neuron's tuning to that cue for that stimulus S,

$$d_y[S] = \max_t d_{y,t}[S]. \quad (12)$$

[Enlarged marker symbols in Fig. 3(e) indicate maxima.] The normalized pre-spike cue histograms with maximal $d_{y,t}$ from Fig. 3(e) are shown in Fig. 3(f) for the three stimulus cues. They represent the final tuning curves, and the corresponding value of d_y represents the tuning measure.

The gain or attenuation in tuning to y (=IPD or BCI) caused by rapid compared to slow fluctuations in a binaural cue, x=IPD, IID, or BCI, was simply measured by the difference in d_y between these two cases (rather than ratio since it is already a logarithmic measure):

$$g_y[x] = d_y[x = \text{fast}] - d_y[x = \text{slow}]. \quad (13)$$

2. Experiment 2: Persistent suppression

For the second experiment, which explicitly examined persistent suppression using the triphasic stimuli (see Sec. II A), a measure of the degree of the persistent component of the suppression was required. Recall that these stimuli consisted of a 20-ms excitatory binaural tone, followed by a 20-ms suppressive binaural tone, followed by the excitatory binaural tone again for 100 ms [e.g., Fig. 4(b)]. The degree of suppression during the second tone was controlled by setting either the IID or the IPD appropriately. During the final (excitatory) tone, any persistent suppression was revealed as a reduced spike-rate relative to the excitatory asymptotic rate [this is evident in the schematic in Fig. 4(a), which would correspond to a case of strong persistent suppression in our data]. The degree of persistent suppression was measured by comparing the mean spike-rates between suppressive and control stimuli during the 30-ms period following the end of the suppressive tone [i.e., between 50 and 80 ms adjusted for neural propagation delay; see Sec. II A and Fig. 4(a)]. A 30-ms period was found to adequately cover the major persistently suppressive period of most neurons. The 30-ms-averaged rates were compared using a ratio between the most and the least heavily suppressed response. To define "most heavily suppressed" and "least heavily suppressed," the response during the 20-ms suppressive tone was used, rather than the 30 ms following it (that was used to measure the degree of *persistent* suppression). This gave the following index:

$$h_{PS}[x] = 10 \log_{10}\left(\frac{\text{mean spike rate of most suppressed}}{\text{mean spike rate of least suppressed}}\right), \quad (14)$$

where x=IPD or IID, depending on which cue was being investigated for persistent suppression. Note that, although Eq. (14) for $h_{PS}[x]$ has some similarities to Eq. (10) for $g_{rate}[x]$, the two quantities are entirely different as they are derived from the response to separate types of stimuli.

III. RESULTS

A. Neural classification

To test the selective filtering hypothesis, we recorded extracellularly from 120 low-frequency neurons in the DNLL of Mongolian gerbils (BF between 0.5 and 3 kHz). This population can conveniently be divided into four groups based on the type of dominant drive they received from each ear. First are neurons with the same type of dominant monaural drive from each ear (n=54): either E/E (excitatory from both sides, 24/52, 44%) or I/I (inhibitory from both sides, 30/52, 56%). For E/E neurons, the excitatory dominant drive was evident through monaural stimulation for some neurons, but for others was only evident as an increase in the spike-rate for binaural compared to monaural stimulation. The inhibition in I/I neurons was observed because these cells always had spontaneous spike-rates (>10 Hz) that were suppressed by a tone of appropriate frequency played to either ear. Second are neurons with the opposite type of drive from opposing ears (n=53): they consisted predominantly (50/53,94%) of I/E neurons, meaning that they experienced a net inhibition from the ipsilateral ear and a net excitation from the contralateral ear, but they also included a small number of E/I neurons (3/53,6%). Excitation was always evident from monaural stimulation for these neurons, while inhibition was evident as a reduced spike-rate when stimulating against a fixed contralateral excitatory input. Third are neurons with monaural input (n=6), all of the 0/E type with excitation from the contralateral ear. Fourth are neurons with complex binaural input (that varied from excitatory to inhibitory as sound pressure level or frequency changed) (n=8). For simplicity, we report data on the first two groups only since they comprise a large majority (88%) of the neurons we recorded from and since they differed in two ways that have important implications for the selective filtering hypothesis (to be described shortly). Neurons in the first group, with the same binaural drive, will be termed X/X neurons, while those in the second group, with opposing binaural drive, will be termed X/Y neurons, where X and Y can stand for either E or I. Our conclusions do not change if the second, third, and fourth groups are included in the following analysis as a single group.

The X/X and X/Y neurons differ in two ways that have important implications for the selective filtering hypothesis. These two distinguishing properties are easily seen in the joint IPD-IID tuning of neurons, as illustrated in Figs. 5(a) (X/X neuron) and 5(b) (X/Y neuron) for representatives from each group. These figures were compiled from the response of the neurons to 160 s of band-limited noise, centered on the neurons BF, with IPDs and IIDs that vary randomly, independently, and slowly [i.e., stimulus S(IPD=s,IID=s,BCI=r), see Table II and Sec. II]. In the plot, for any joint IPD-IID bin, the grayscale shows a histogram of the spike-rate averaged over all 10-ms intervals of the stimulus

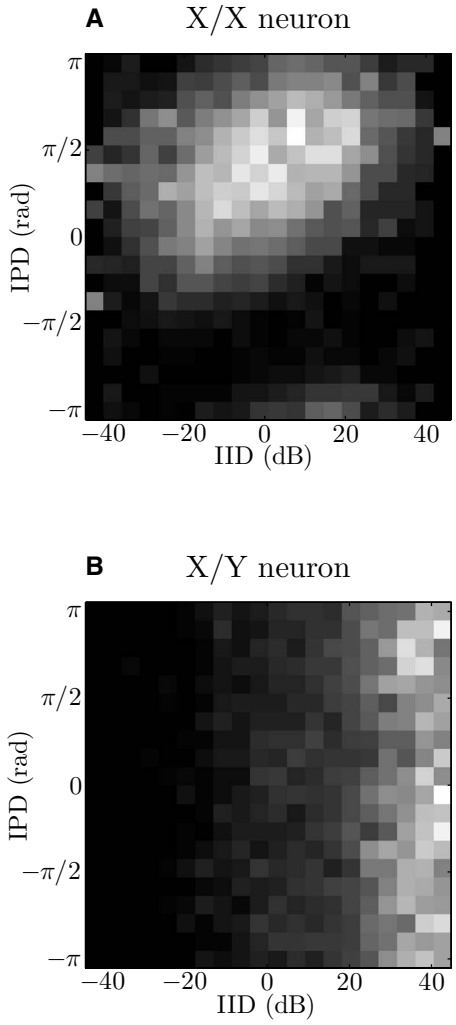


FIG. 5. Characteristic joint IPD-IID tuning of X/X and X/Y neurons. (A) An X/X neuron. (B) An X/Y neuron. In each plot the grayscale indicates the mean spike-rate as a function of IPD and IID, with white being the greatest [equivalent to (A) 33 Hz or (B) 26 Hz] and black indicating zero spike-rate.

during which that particular IPD-IID value prevailed. Bin sizes were $\pi/10$ rad for IPD and 4.5 dB for IID.

The first, defining, property of the two groups is that they differed in their type of sensitivity to IID. The X/X neuron in Fig. 5(a) showed a peaked spike-rate at a central IID around zero, whereas the X/Y neuron in Fig. 5(b) had a relatively low spike-rate at this IID value compared to its peak rate at a peripheral IID (in this case=45 dB, which was the maximum value tested; a positive value indicates that the sound is louder at the contralateral ear). All X/X neurons had a peak rate at an IID between −10 and 10 dB, while X/Y neurons had peaks outside this range, typically at the maximum or minimum tested (±45 dB).

The second distinguishing property is that the two groups differed in their sensitivity to IPD. The X/X neuron in Fig. 5(a) was sensitive to IPD, as shown by the modulation in spike-rate with this parameter, whereas the X/Y neuron in Fig. 5(b) showed no significant variation in spike-rate as IPD varied. In our sample, the spike-rate of X/X neurons was almost always sensitive to IPD (51/54,94%), while the great majority (40/53,75%) of X/Y neurons were insensitive to this parameter (leaving 25% that were sensitive). [A neuron was labeled IPD sensitive if its spike-rate at best IPD was more than 50% greater than its spike-rate at worst IPD (where best/worst IPD gave maximal/minimal rate averaged across IID).]

The significance of these two observations for the selective filtering hypothesis can be understood in terms of the differing importance of IPD and IID for the localization of low-frequency sounds. First, a low-frequency neuron that is selectively filtering out spurious localization cues should, as a fundamental requirement, be sensitive to IPD because IPD is the dominant low-frequency localization cue. Second, the same neuron should also exhibit a non-negligible spike-rate for small IIDs (given an appropriately excitatory IPD) because the IID is always small at low frequencies, regardless of location, for a (far-field) sound source uncorrupted by other sources. For these two reasons, we hypothesized that, of the two groups of low-frequency neurons, the X/X neurons should exhibit selective filtering by rapidly fluctuating IPDs and IIDs, whereas X/Y neurons should not. It is important to note that we are still hypothesizing the suppression of response in X/X neurons by rapidly fluctuating IIDs. However, we are not considering the suppression of IID tuning in low-frequency X/Y neurons because IID is not a dominant localization cue at low frequencies. This is not to rule out the possibility that IID tuning is suppressed by rapidly fluctuating IIDs in high-frequency neurons: however, we do not present any data on this topic.

B. Experiment 1: Selective filtering

To test the selective filtering hypothesis directly, we presented BF narrowband noise in which the binaural cues IPD, IID, and BCI varied either rapidly or slowly (see Sec. II). According to the hypothesis, we expected to see an attenuation of response for rapid compared to slow stimuli only in the case of IPD and IID, but not for BCI.

To quantify the effect of rapid compared to slow fluctuations of some binaural cue, x=IPD, IID, or BCI, we used three complementary measures. The first measure was a simple comparison of the stimulus-averaged spike-rates evoked by stimuli with rapid compared to slow fluctuations. This was expressed as a gain, $g_{\text{rate}}[x]$, calculated as the ratio of rates, with the result converted into a decibel scale [see Sec. II, Eq. (10)]. Negative values indicate attenuation for rapid compared to slow fluctuations, whereas positive values indicate a gain, with $g_{\text{rate}}[x]=3$ dB nearly equivalent to a doubling of spike-rate.

The stimulus-averaged spike-rate gives information about the overall strength of the response to the stimulus, but does not provide information about the tuning of that response with respect to a particular binaural cue, such as y =IPD or BCI. In order to compare the tuning response of a neuron under different stimulus conditions, two more measures of another type were used (one for IPD and another for BCI; see Sec. II and Fig. 3). This first involved estimating the tuning curves for a binaural cue from the response to the narrowband stimuli. These curves may be interpreted as giving the mean spike-rate evoked by a particular cue value, normalized so as to give a probability function of unit sum.

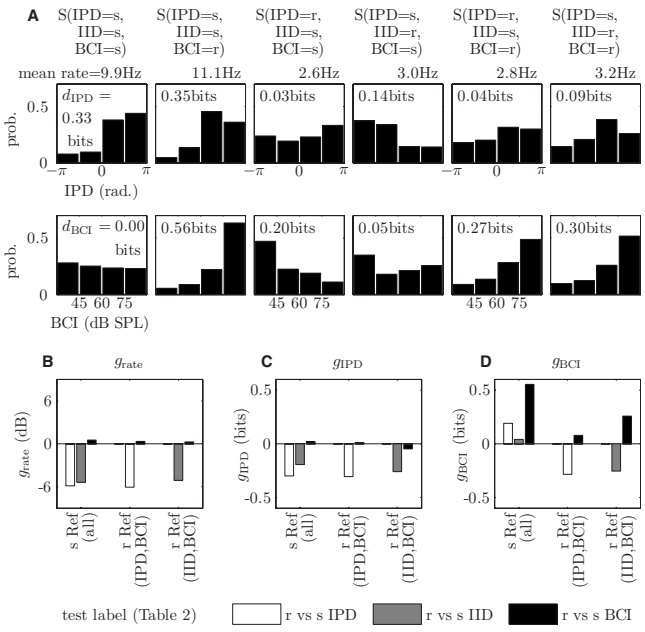


FIG. 6. Illustration of the main measures of selective filtering used in this study: g_{rate}, g_{IPD}, and g_{BCI}. (A) Maximal normalized spike-triggered histograms (NSSHs) for IPD (top row) and BCI (bottom row) for the six stimuli types used in this study (labeled; see Table I). The stimulus-averaged spike-rates, d_{IPD}, and d_{BCI}, for each stimulus are indicated on each plot. [(B)–(D)] The measures (ordinate) used in this study to quantify the effect of rapid vs slow cue fluctuations are plotted for the three test sets listed in Table II (abscissa): (B) g_{rate} [Eq. (10)], (C) g_{IPD}, and (D) g_{BCI} [both Eq. (13)]. Bars comparing fast vs slow fluctuations in IPD, BCI, and IID are coded in three shades of gray, as shown in the legend.

The normalization allowed us to measure tuning via the curve's divergence from the uniform (untuned) curve using the Kullback–Leibler divergence, d_y [Sec. II, Eq. (11)] (Kullback, 1959). The Kullback–Leibler divergence is derived from information theory and is measured in bits. d_y has a minimum of 0 bit for the uniform distribution (i.e., a flat curve, no tuning) and increases to a maximum of 2 bits (here) when all the probability is concentrated at a single value (bin) (i.e., a peaked curve, maximal tuning). The gain or attenuation, $g_y[x]$, in tuning to y (=IPD or BCI) caused by rapid compared to slow fluctuations in a binaural cue x (=IPD, IID, or BCI) was measured by the difference in the divergence between these two cases [see Sec. II, Eq. (13)]. Note that the tuning gain g_y is explicitly independent of the mean spike-rate since it depends on a probability measure that is independent of spike-rate: Eq. (11). Consequently, it is a complementary measure to the rate gain g_{rate}.

We were interested in the tuning of low-frequency neurons to the dominant localization cue (y=IPD) and to amplitude modulation (y=BCI), so we used g_{rate}, g_{IPD}, and g_{BCI} as our three measures to quantify the effect of rapid compared to slow fluctuations.

An illustration of the neural response measures, described above, is given in Fig. 6 for one particular X/X neuron. Figure 6(a) shows the tuning curves for binaural cues y=IPD (top row) and BCI (bottom row) for all six stimuli listed in Table I (columns 1–6, as labeled). [That is, the maximal normalized pre-spike cue histograms, as described in Sec. II, are plotted in Fig. 6(a)]. In these plots, tuning is evident as a favoring of some bins over others. The value of the tuning measure d_y is marked on each histogram. The positive relationship between the degree of tuning apparent in the histograms and the corresponding value of d_y is evident. Also marked in Fig. 6(a), under each stimulus heading, is the stimulus-averaged spike-rate. Notice that the two leftmost stimuli in Fig. 6(a) involve slowly fluctuating IPD and IID, whereas the four rightmost stimuli have either IPD or IID (or both) fluctuating rapidly. Generally, the stimuli with slowly fluctuating IPD and IID had large mean spike-rates and strong IPD tuning (i.e., large d_{IPD}) compared to the stimuli with rapidly fluctuating IPD or IID. For this particular neuron, no such pattern was evident in BCI tuning.

These types of comparisons allowed us to directly test the selective filtering hypothesis. They were made more specific, and quantified, by using the comparative response measures g_{rate}, g_{IPD}, and g_{BCI}, according to the particular pairs of stimuli listed in Table II. These pairs of stimuli differed only in the rates of fluctuation of a single specified cue, x=IPD, IID, or BCI: one stimulus had rapid fluctuations; the other slow (Sec. II describes the choice of these stimuli in greater detail). The results are shown in Figs. 6(b)–6(d) for mean rate (g_{rate}), IPD tuning (g_{IPD}), and BCI tuning (g_{BCI}), respectively. In each plot, the height of the bar indicates the gain in the appropriate response measure (ordinate) for rapid compared to slow fluctuations in some cue x. Bars are clustered into three groups corresponding to the three test sets "s Ref (all)," "r Ref (IPD, BCI)," and "r Ref (IID, BCI)" listed in Table II (abscissa). These test sets differed in which stimulus served as a reference when making comparisons across IPD, IID, and BCI fluctuations (see Sec. II for more detail). According to the selective filtering hypothesis, comparisons with rapid vs slow fluctuations in x=IPD (white bars) or IID (gray bars) were predicted to show a suppression of neural response ($g<0$), while comparisons with x=BCI (black bars) were predicted to show no neural response suppression ($g\geq 0$).

For the neuron referred to in Fig. 6, rapid compared to slow fluctuations of either IPD or IID produced large negative values of g_{rate} [Fig. 6(b)] and g_{IPD} [Fig. 6(c)] across all three test sets, indicating that both mean spike-rate and IPD tuning were attenuated by the rapid fluctuations. By comparison, rapid compared to slow fluctuations in BCI (black bars) caused neither a marked attenuation nor a marked gain in the mean rate or the IPD tuning of this neuron. For this particular X/X neuron, tuning to BCI [Fig. 6(d)] was not consistently affected by rapid compared to slow fluctuations in either IPD or IID: one test showed an attenuation of response (the "r Ref" test set), whereas the other set showed a gain [the "s Ref (all)" test set]. This pattern of BCI tuning was not observed in the majority of X/X neurons. Tuning to BCI was generally enhanced by rapid compared to slow fluctuations in BCI for this neuron, as shown by consistently positive black bars in Fig. 6(d).

A summary of population data for $g_y[x]$, y=rate, IPD, or BCI, is shown in the three rows of Fig. 7 (top to bottom, respectively, as labeled). The three columns of the figure correspond to an analysis of rapid compared to slow fluctuations of x=IPD, IID, or BCI (left to right, respectively, as

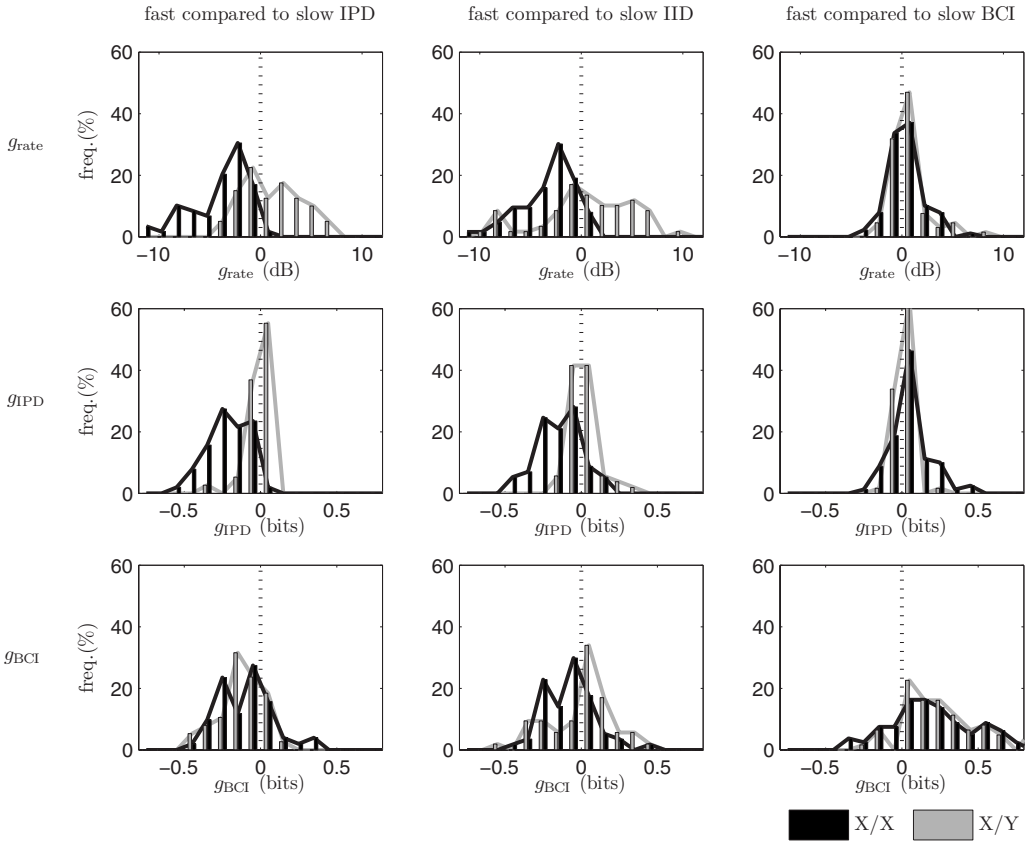


FIG. 7. Testing the selective filtering hypothesis: population data on the effect of rapid vs slow fluctuations in a binaural cue (IPD, IID, and BCI) upon neural response. Each plot is a histogram depicting the distribution of a comparative response measure, $g_y[x]$, across the population of X/X (black) and X/Y (gray) neurons. $g_y[x]$ measures the gain ($g_y[x]>0$) or attenuation ($g_y[x]<0$) in some neural response measure for rapid compared to slow fluctuations of x=IPD, IID, or BCI (columns 1–3, left to right, respectively, as labeled). Three different neural response measures were used: stimulus-averaged spike-rate (y=rate), rate-independent IPD tuning (y=IPD), or rate-independent BCI tuning (y=BCI) (rows 1–3, top to bottom, respectively, as labeled).

labeled). Each plot is a histogram depicting the distribution of $g_y[x]$ across the population of X/X (black) and X/Y (gray) neurons. (For simplicity, results have been pooled across all comparisons involving rapid compared to slow fluctuations in x, regardless of whether the accompanying fluctuations in other binaural parameters were either both fast or both slow. That is, in Table II, results have been pooled over columns to give the following three groups: IPD comparisons, IID comparisons, and BCI comparisons. An analysis that does not make this simplification gives very similar conclusions.)

Figure 7 shows that the great majority of X/X neuronal responses (72%–98%, depending on the condition and measure) were attenuated ($g_y<0$) by rapid compared to slow IPD or IID fluctuations (columns 1 and 2) and that the attenuation was often strong. This was true whether attenuation was measured by rate, IPD tuning, or BCI tuning (rate: row 1, >92% with $g_{\text{rate}}<0$ dB; IPD: row 2, >86% with $g_{\text{IPD}}<0$ bit); BCI: row 3, >72% with $g_{\text{BCI}}<0$ bit). In contrast, rapid compared to slow BCI fluctuations (column 3) caused no general attenuation of response across the population: the distribution of g_y was either clustered tightly around 0 or lay mostly over positive values (rate: row 1, >71% with $|g_{\text{rate}}|<1.5$ dB; IPD: row 2, >65% with $|g_{\text{IPD}}|<0.1$ bit; BCI: row 3, >77% with $g_{\text{BCI}}>0$ bit).

By comparison, X/Y neuronal responses showed no tendency at a population level to be attenuated by rapid compared to slow fluctuations in either IPD or IID (columns 1 and 2). [The exception was g_{BCI}[IPD] (bottom, left plot) in which there was an overall attenuation that was similar to that of the X/X neurons.] Generally, under rapidly fluctuating IPD and IID conditions, the difference in the distribution of X/X and X/Y neurons was most pronounced at comparatively large values of attenuation of rate ($g_{\text{rate}}<-3$ dB) and attenuation of IPD tuning ($g_{\text{IPD}}<-0.2$ bit), with X/X neurons predominant in these regions. Again, in contrast to rapid compared to slow BCI fluctuations (column 3), these difference between X/X and X/Y neurons were not apparent, rather the distributions for X/X and X/Y neurons were similar.

In summary, X/X neurons showed a suppression of mean spike-rate and a diminished IPD and BCI tuning for fast compared to slow fluctuations in either IPD or IID. They did not show this pattern of behavior for fast compared to slow BCI fluctuations. In contrast, X/Y neurons only showed attenuation in only one of nine cases: for rapidly fluctuating IPDs when measured using g_{BCI}.

C. Experiment 2: Persistent suppression

To test for the presence of persistent suppression, we identified suppressive subfields in the joint IPD-IID tuning of 32 X/X and 32 X/Y neurons. These subfields corresponded to the relatively dark areas of low spike-rate in Figs. 5(a) and 5(b); for example, the entire region corresponding to nega-

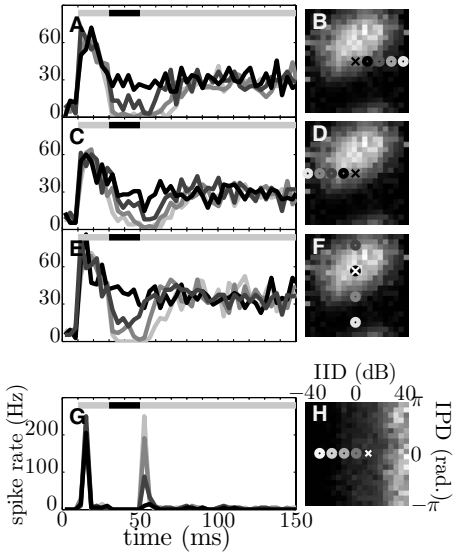


FIG. 8. The presence of persistent suppression in an X/X neuron and absence in an X/Y neuron. The presence of persistent suppression was revealed by presenting a 20-ms BF probe tone with an IID or IPD from a suppressive subfield of the neuron in between two temporally flanking reference tones from an excitatory part of the subfield. [(A)–(F)] X/X neuron. (A) Spiking rate as a function of peri-stimulus time (averaged over 3.4-ms bins) for four different choices of suppressive IID (grayscale coded) as shown by matching grayscale circles in (B). (B) is the joint IPD-IID tuning of the neuron (see Fig. 5 for further details). The cross in (B) indicates the IPD-IID chosen for the excitatory part of the stimulus. The bar under the time-axis in (A) indicates the excitatory (gray) and suppressive (black) portions of the stimulus, shifted by 10 ms to correct for the neural propagation delay. (C) and (D) and (E) and (F) are similar to the pair (A) and (B) except that they use suppressive tones from the other two suppressive subfields of the neuron, i.e., negative IIDs and IPDs around −2 rad respectively. [(G) and (H)] The same protocol used in (A)–(F) is applied to an X/Y neuron.

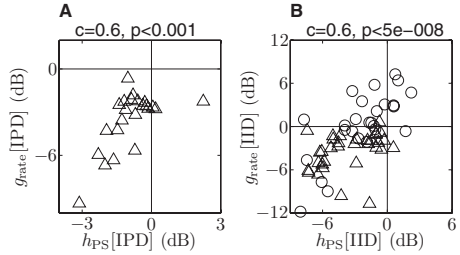


FIG. 9. The correlation between the degree of selective filtering of mean spike-rate and the degree of persistent suppression. (A) Scatter plot, over the neural population, of the degree of attenuation/gain for rapid compared to slow IPD fluctuations (g_{rate}[IPD], ordinate) vs maximal degree of persistent suppression evoked by a suppressive IPD (h_{PS}[IPD], abscissa). (B) Scatter plot as per (A) but with IID instead of IPD throughout. Triangles correspond to X/X neurons and circles X/Y neurons. The correlation coefficient (c) and significance (p) for the pooled data are shown on top of each plot.

tive IIDs in Fig. 5(b), and three regions in Fig. 5(a) corresponding to large negative IIDs, large positive IIDs, and IPDs around −2 rad. We attempted to elicit a persistent suppression by presenting a 20-ms BF tone, with IPD and IID chosen from a neuron's suppressive subfield. In order to observe persistent suppression it was necessary to also provide an excitatory stimulus that temporally flanked the suppressive tone so that any suppression would be revealed a reduction in the spike-rate from that expected from the excitatory stimulus. This was achieved by playing a BF tone with excitatory IPD and IID for 20 ms immediately before the suppressive tone and for 100 ms immediately following it (see Fig. 4 and Sec. II). This showed the time course of suppression as a reduction in spike-rate compared to the asymptotic level of excitation.

For some neurons, such as the one for which data are presented in the top three panels of Fig. 8, this protocol revealed a component of the suppression that lasted beyond the 20-ms duration of the suppressive part of the stimulus. [Note that this is for the same X/X neuron shown in Fig. 5(a).] Figure 8(a) shows four overlaid traces of peri-stimulus time histograms obtained by applying this protocol with four increasingly suppressive IIDs=10, 20, 30, and 40 dB (and IPD=0 rad) as coded by the four gray-matched circles in Fig. 8(b). Each trace shows an initial period of relatively low background activity, approximately 10 ms in duration corresponding to the neural propagation delay, followed by 20 ms of elevated rate evoked by the excitatory tone (IID=0 dB), followed by a period of suppression evoked by the suppressive tone, and finally a period of elevated rate due to the final excitatory tone [also IID=0 dB as indicated by the cross in Fig. 8(b)]. For an IID=20, 30, and 40 dB, the section of the trace following the end of the suppressive tone (at 50 ms, corrected for propagation delay) reveals a persistence of suppression for around 10 ms for the least suppressive IID =20 dB and for more than 20 ms in the case of the most suppressive IID=40 dB. This persistence is in contrast to the rapid change (<2 ms) in spike-rate observed at the beginning of the first excitatory stimulus (at 10 ms) and the change from excitatory to suppressive stimulus (at 30 ms). For this particular neuron, we were able to record for long enough to find that persistent suppression could also be evoked from the other two suppressive subfields, which involved either large negative IIDs [Figs. 8(c) and 8(d)] or suppressive IPDs [Figs. 8(e) and 8(f)].

To quantify the persistent suppression, the mean spike-rates for a 30-ms period following the end of the suppressive tone (i.e., between 50 and 80 ms) were compared using a ratio between the most and the least heavily suppressed (see Sec. II). Expressed in decibels, this gave an index, $h_{\text{PS}}[x]$, in which negative values indicated suppression and positive values indicated facilitation, and x=IPD or IID depending on which binaural cue was examined. The examples in Figs. 8(a), 8(c), and 8(e) gave h_{PS}[IID]=−4.9, −3.8 dB and h_{PS}[IPD]=−2.3 dB, respectively. Overall, of the 24 X/X neurons tested with both suppressive IIDs and suppressive IPDs, 19 (79%) showed some persistence of suppression in both cases (h_{PS}[IID]<0 dB and h_{PS}[IPD]<0 dB), though the persistence was generally less in the case of IPDs (see also Fig. 9 for an indication of the relative magnitude of h_{PS}[IID] and h_{PS}[IPD]).

An example in which facilitation, instead of persistent suppression, was observed is shown on the bottom plot of Fig. 8 [which is for the same X/Y neuron, as shown in Fig. 5(b)]. This neuron had an h_{PS}[IID]=9.3 dB, indicating that the tone with IID that was deepest in the suppressive subfield (IID=−30 dB) actually led to a much larger response in the 30 ms following the suppressive part of the stimulus than the tone that was only on the edge of the suppressive subfield

(IID=0 dB). This is because the neuron showed a rebound effect at the onset of a rapid change from negative to positive IIDs that was more pronounced the more negative the first IID became [Fig. 8(g)]. This rebound-onset type of response was typical of neurons that exhibited a facilitative effect.

Most X/X neurons (>84%) showed some degree of suppression over the 30 ms following the suppressive part of the stimulus, as measured by h_{PS}. For suppressive IIDs, 56% (18/32) showed a strong suppression (h_{PS}[IID]<−3 dB, i.e., more than 50% suppression), 41% (13/32) showed a mild suppression (−3<h_{PS}[IID]<0 dB), and only 3% (1/32) showed facilitation (h_{PS}[IID]>0 dB). For suppressive IPDs, 4% (1/25) showed a strong suppression, 80% (20/25) showed a mild suppression, and 16% (4/32) showed facilitation. A smaller, but nonetheless major, portion of X/Y neurons also showed either strong (34.5%, 11/32) or mild (34.5%, 11/32) suppression for IIDs, but in contrast to X/X neurons, there was also a large portion (31%, 10/32) that exhibited facilitation. The lack of clear IPD tuning in most X/Y neurons precluded investigating whether IPD had a persistently suppressive effect in these neurons.

To further assess the degree to which persistent suppression might underlie the selective attenuation of mean spike-rate by rapidly varying IPDs and IIDs, we looked for a correlation between the two phenomena. This was done by plotting g_{rate}[IPD] or g_{rate}[IID] against the strongest h_{PS}[IPD] or h_{PS}[IID], respectively. Intuitively, this plots the degree of attenuation brought about by rapid compared to slow IPD or IID fluctuations against the degree of persistent suppression evoked by the maximally suppressive IPD or IID, respectively. This is shown in Fig. 9 for X/X (triangles) and X/Y (circles) neurons. For IPD, there was a significant ($p<0.001$, two tailed t-statistic) correlation between h_{PS}[IPD] and g_{rate}[IPD] (correlation coefficient=0.6) for X/X neurons [Fig. 9(a)]. [X/Y neurons are not shown in Fig. 9(a) because they generally lacked IPD sensitivity, thus making it meaningless to test for persistent suppression from a suppressive IPD.] For IID, there was a significant correlation between h_{PS}[IID] and g_{rate}[IID] for X/X neurons (correlation coefficient=0.4, $p<0.05$), X/Y neurons (correlation coefficient=0.7, $p<10^{-4}$), and both X/Y and X/Y neurons taken together (correlation coefficient=0.6, $p<10^{-7}$).

Note that only negative values of h_{PS}[IID] could plausibly correspond to a true persistent inhibition; positive values of h_{PS}[IID] were associated with an onset/offset behavior in X/Y neurons, in which persistent inhibition was apparently absent [e.g., Fig. 8(g)]. X/Y neurons were a heterogeneous group, with many showing neither persistent suppression nor selective filtering, while other X/Y neurons showed both or one but not the other. Nonetheless a positive and significant correlation exists between h_{PS}[IID] and g_{rate}[IID] for X/Y (as well as X/X) neurons.

The effect of persistent suppression on the rate-independent aspects of IPD and BCI tuning, as measured by g_{IPD} and g_{BCI}, respectively, is not as straightforward to predict as the effect on g_{rate}. Indeed, there were no clear correlations between persistent suppression and these two measures in our data.

IV. DISCUSSION

A. General discussion

In this study we have considered the hypothesis that a selective filtering acts in the DNLL whereby rapid compared to slow fluctuations in IPD and IID lead to an attenuation of neural response, whereas rapid compared to slow BCI fluctuations do not. Functionally, this would result in the suppression of neural response to spurious, but not valid binaural localization cues (see Introduction). Using stimuli that differed primarily in the rate at which IPD, IID, or BCI fluctuated, we identified a population of low-frequency neurons in the gerbil DNLL with matched binaural drive type (X/X neurons) that largely satisfied this hypothesis for three independent response measures: mean spike-rate, (rate-independent) IPD tuning, and, to a lesser extent, (rate-independent) BCI tuning (Fig. 7). A second population of low-frequency DNLL neurons with opposing binaural drive type (X/Y neurons) exhibited responses that were generally inconsistent with the selective filtering hypothesis. This difference in response emerged as a secondary distinguishing feature of X/X and X/Y neurons since they were primarily distinguished on the basis of their joint IPD-IID tuning (Fig. 5). By definition, X/X neurons were all sensitive to IPD and had maximal spike-rates at IIDs close to 0 dB. In contrast, X/Y neurons usually showed no IPD tuning and had maximal spike-rates at large (positive or sometimes negative) IIDs.

For a single sound source at low frequencies, IPDs are the primary azimuthal localization cue and IIDs are usually small. This suggests that X/X neurons are well suited to conveying information about the location of a single low-frequency sound source, while X/Y neurons are not. Here, as summarized in the previous paragraph, we have provided evidence that X/X neurons are also well suited to filtering out spurious localization cues as when multiple sound sources are present.

On the other hand, the role of X/Y neurons at low frequencies is mysterious because, at these frequencies, they are often tuned to large IIDs outside the range obtainable from a single sound source and because of the low importance of IID as localization cue at these frequencies. On a speculative note, from the current perspective, these neurons may play a role in detecting the large, spurious IIDs caused by interfering sound sources. They may even play a role in suppressing the response of other neurons in such circumstances, given that they are GABA-ergic.

One criticism of the selective filtering hypothesis is that it may not be necessary to have a neural filter that lets rapid BCI fluctuations pass but suppresses rapid IPD and IID fluctuations originating in the superior olive. This is because the BCI fluctuations may be passed through a parallel pathway in the auditory system. While it is likely that such pathways exist, this argument supposes that the only reason to filter out rapid IPD and IID fluctuations selectively is to ensure that the rapid BCI fluctuations still reach higher processing stages of the brain. However, this view overlooks the point that rapid fluctuations of IPD or IID convey something very different about the auditory scene than do rapid fluctuations in BCI. As summarized in the Introduction, the former indicate

the interference of multiple sound sources and the degradation of localization information, while the latter may provide important information about a single sound source. As a result, it seems likely to be important for the brain to be able to distinguish between the two kinds of rapid fluctuations. It is not clear how this could be done on the basis of the output of a superior olive neuron, which is typically modulated by both BCI and IPD or IID (or both); the source of the modulation is confounded. The selective filtering hypothesis proposes a way in which this distinction might be made: rapid IPD and IID fluctuations lead to suppressed response, while rapid BCI fluctuations do not. This scheme has the potential advantage that information about location and amplitude modulation of a sound source can be conveyed in a single pathway when interfering sources do not impinge.

In the second part of Sec. III, we examined the hypothesis that the observed selective filtering of mean spike-rate (but not IPD tuning) is due to a so-called persistent suppression. Such a persistent suppression was revealed in some neurons as an ongoing (5–30-ms) suppression of activity following the end of the suppressive part of a stimulus (Fig. 8). Both suppressive IPDs and IIDs were capable of inducing the persistent component of the suppression in X/X neurons. A GABA-ergic persistent inhibition that is also induced by inhibitory IIDs has been observed previously in the DNLL using a stimulation protocol that has similarities to the one used here (Yang and Pollak, 1994a, 1994b; Burger and Pollak, 2001). Burger and Pollack (2001) and Pecka *et al.* (2007) presented evidence that this arises from the contralateral DNLL. Thus, it seems probable that the persistent suppression observed in this study is a GABA-ergic inhibition from the contralateral DNLL.

We argued that such circuitry, with its complementary tuning of excitation and persistent inhibition (as a function of IID and IPD), would lead to a selective filtering of mean spike-rate in persistently suppressed neurons of the DNLL (see Fig. 2). In support of this hypothesis, we found that the majority of X/X neurons showed a degree of selective filtering and were persistently suppressed by both suppressive IPDs and IIDs. Further, the degree of selective filtering of rate by rapidly fluctuating IPDs and IIDs was positively and significantly correlated with the degree of persistent suppression induced by suppressive IPDs and IIDs, respectively (Fig. 9). By contrast, X/Y neurons exhibited a more heterogeneous response: a large portion showed neither selective filtering (approximately 50%) nor persistent suppression (31%), while others showed either or both. Despite this, a positive correlation was also found between persistent suppression and selective filtering in X/Y neurons. Thus, the data on both X/X and X/Y neurons are consistent with the idea that persistent inhibition in the DNLL plays a role in selective filtering of mean spike-rate by rapidly fluctuating IPDs and IIDs. For X/X neurons this appears to be true of most neurons, while for X/Y neurons it only applies to a (minor) subpopulation.

It must be acknowledged that the data relating selective filtering to persistent suppression are purely correlational and so cannot establish the cause of the selective filtering. Further, the correlations themselves are not especially strong. This raises the possibility that other mechanisms, perhaps earlier in the auditory pathway, are responsible for the observed selective filtering. Further experiments in which, for example, the degree of selective filtering in the superior olivary complex was measured, activity in the contralateral DNLL was suppressed, or GABA-ergic inhibition in the DNLL was blocked, would help to clarify the mechanism(s) involved. Indeed, we attempted the latter experiment, but were unable to obtain consistent results from the iontophoretic application of synaptic blockers (even in the inferior colliculus) during the time available.

The stimuli used to test for and measure persistent suppression involved setting one binaural parameter to zero and varying the other, i.e., setting IID=0 dB and varying IPD or vice versa (see Experiment 2 in Sec. II A). A potential problem with this approach is that fixing one parameter to zero does not take into account a particular neuron's tuning since it may bear no special relationship to the set of values with IPD=0 rad and IID=0 dB. Consequently, this may have been a source of undesirable variability in the measure of persistent suppression, $h_{PS}[x]$ [Eq. (14)].

An alternative approach would have been to define the fixed values for IID and IPD to be those giving maximal response (IID_{max} and IID_{max}, respectively) and to use an excitatory reference tone with $IID_{ex}=IID_{max}$ and $IID_{ex}=IID_{max}$. [This can be visualized as having the cross, representing the reference tone, on the whitest square in Figs. 8(b) and 8(h), then circles, representing the suppressive probe tone, at values that are either horizontally (IID) or vertically (IPD) displaced from this.] This would have taken account of neural tuning and would have had the additional advantage of having a single excitatory reference tone for variations in both IID and IPD. Consequently, this may have yielded measures of persistent suppression that were more consistent across neurons and/or between IID and IPD evoked suppression. This in turn may have led to clearer results when examining the correlation between selective filtering and persistent suppression (Fig. 9).

However, a drawback of this approach is that there was typically no well-defined IPD giving maximal response for X/Y neurons since most of these neurons were insensitive to variation in IPD. As a result, this would have required choosing the fixed IPD for X/Y neurons according to other criteria than that used for X/X neurons. Furthermore, maximal IIDs for X/Y neurons are typically large in magnitude, whereas they are typically small in magnitude for X/X neurons. This would have led to very different excitatory reference IIDs for X/X vs X/Y neurons. For these reasons, subsequent measures of persistent suppression may still not be comparable between X/X and X/Y neurons. As it was, choosing a fixed IPD to be 0 rad and varying the suppressive IID typically led to excitatory reference tones that were roughly half maximal activation, which was consistent for both X/X and X/Y neurons. This is because the peaks of IPD tuning curves are typically displaced toward positive IPDs and have submaximal response at 0 rad (Brand *et al.*, 2002) Similarly, setting the fixed IID to 0 dB when varying the suppressive IPD, which was only done for X/X neurons, should have given fairly consistent results across neurons. This is because

the peak activation for X/X neurons always lay between −10 and 10 dB. More caution needs to be taken, however, when comparing measures of persistent suppression evoked by IIDs vs IPDs within the X/X population. The reason is that the excitatory reference tone for IIDs typically gave a submaximal response, but for IPDs was chosen to give near-maximal response. Consequently, it may be incorrect to interpret the generally smaller values of h_{PS}[IPD] compared to h_{PS}[IPD], indicating that persistent inhibition is weaker when evoked by IPDs than when evoked by IIDs.

B. Relationship to previous studies

In a study in the barn owl, Keller and Takahashi (2005) examined an alternative selective filtering hypothesis to the one examined here. This involved neuronal integration of information about ITDs and IIDs across tonotopic frequency bands (see Introduction for more details). While they provided data supporting this hypothesis, their model cannot explain the results presented here because we used narrowband stimuli that precluded any integration across frequency.

The substantial portion of low-BF EI or IE neurons in the DNLL, observed in this and other studies (Sieveke *et al.* 2006), may appear puzzling given that IIDs from a single sound source are typically small at these frequencies. However, when multiple sound sources are present and the signal becomes corrupted, IIDs need not be small at low frequencies (see Fig. 1). One possible role for these neurons is that they detect the large excursions in IID that are indicative of such a corrupted signal. Indeed, such a response property would make them an ideal candidate for the source of the persistent inhibition in the selective filtering hypothesized here.

Persistent inhibition in the DNLL has also been hypothesized to play a role in echo suppression (Yang and Pollak, 1994a, 1994b; Burger and Pollak, 2001; Pecka *et al.*, 2007): e.g., in the psychophysically established precedence effect (Zurek, 1987; Blauert, 1997; Litovsky *et al.*, 1999). In this effect, information about the direction of a sound is suppressed when it trails a first sound by between 2 and 20 ms, but other information, such as about intensity, is retained (Freyman *et al.*, 1998). A role for persistent inhibition in this effect has similarities to the role in selective filtering of spurious localization cues that is proposed here. Indeed, the two proposed roles are mutually consistent.

The results presented here may also be relevant to another psychophysical phenomenon, the so-called "binaural sluggishness" of the auditory system (Grantham and Wightman, 1978; Grantham, 1982; Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998; Culling and Colburn, 2000; Bernstein *et al.*, 2001; Boehnke *et al.*, 2002). This is manifested as a severely limited ability of listeners to follow changes in location, through ITDs and IIDs, or changes in interaural correlation, compared to their ability to follow changes in amplitude. Indeed, subjects are insensitive to location for modulation rates exceeding 20 Hz for ITD, IID, or interaural correlation (although see Sieveke *et al.*, 2007), whereas rates of amplitude modulation greater than 300 Hz are readily detected. Again, this has qualitative similarities to the selective filtering demonstrated in our results. However, even the slow stimuli used in the present study had significant spectral energy up to rates of 20 Hz for IPD and IID fluctuations. Consequently these stimuli have rates of fluctuation in IPD and IID that lie above the range that human listeners are sensitive to. Despite this, our results show that there was still significant tuning to IPD in X/X neurons at these rates of fluctuation (i.e., up to 20 Hz) when compared to the faster rates. This is consistent with other recent studies (Joris *et al.*, 2006; Siveke *et al.*, 2007) in the cat inferior colliculus and gerbil DNLL, respectively, which specifically addressed the issue of binaural sluggishness. We concur with the conclusion of Joris *et al.* (2006) that further filtering at higher levels in the auditory pathway is required to fully explain binaural sluggishness, although the action of persistent suppression studied here may constitute an early step in the processing underlying binaural sluggishness.

Another aspect of the psychophysics of interaurally decorrelated sounds, which should be mentioned here, is binaural unmasking (Hirsh, 1948; Durlach and Colburn, 1978; Culling *et al.*, 2003; Ackeroyd, 2004; Culling, 2008; Litovsky and McAlpine, 2009 for a review of neural correlates). In this phenomenon, listeners obtain an advantage in detecting a comparatively weak binaural signal amidst binaural noise when the signal has substantially different binaural cues to the noise compared to when it has the same binaural cues. In frequency channels in which the signal is sufficiently strong, this produces a degree of interaural decorrelation (via the mechanism described in Fig. 1), which listeners use to detect the signal. In a series of experiments, Palmer *et al.* (1999) found evidence for a neural correlate of this effect in the inferior colliculus (Caird *et al.*, 1991; McAlpine *et al.*, 1996; Jiang *et al.*, 1997a, 1997b). The correlate was most clearly demonstrated in the final two of these studies, in which they found that the threshold for detection of signals with the same binaural cues was greater than the thresholds for detection of signals with different binaural cues: a clear correlate of the psychophysical effect. Interestingly, the lowest thresholds for detection of signals with different binaural cues occurred in one population of neurons due to a decrease in spike-rate, whereas the lowest thresholds for detection of signals with the same binaural cues occurred in a second population of neurons due to an increase in spike-rate. In a subsequent paper, they argued that the main aspects of these results could be explained on the basis of the standard cross-correlation model of ITD sensitivity (Palmer *et al.*, 1999).

It is possible that the responses of X/X neurons reported here are involved in binaural unmasking since they relate to the detection of interaural decorrelation. However, attempts to relate the present results to those mentioned above are not straightforward. When comparing the effect of sound that is interaurally decorrelated vs correlated, one would expect, on the basis of the present results, an overall disinhibitory effect of DNLL X/X neurons of their (putative) targets in the inferior colliculus (since the DNLL is GABA-ergic). Thus one might expect an increase in spike-rate with signal level for these target neurons when presenting stimuli in which the signal had different binaural cues to the noise (such as used

by Jiang *et al.* 1997a, 1997b). This would appear to be at odds with the idea that X/X neurons in the DNLL play a role in binaural unmasking via the neural correlate described by Jiang *et al.* (1997a, 1997b). This is because they found that the lowest detection thresholds were due to a decrease, rather than increase, in spike-rate. However, it is difficult to directly relate the present results to those described by Jiang *et al.* (1997a, 1997b), as we did not use a stimulus paradigm of a signal added to noise, as was used in those studies. For example, in the studies of Jiang *et al.* (1997a, 1997b) signal detection involved the comparison of responses from noise alone, with IPD=0 or π, with those from signal plus noise, in which IPD would have fluctuated. In contrast, the present study involved comparisons in which both stimuli had fluctuating IPDs, with the difference being the rate of fluctuation. It is, therefore, difficult to relate changes in spike-rate in one stimulus paradigm to those in another.

V. CONCLUSION

The present results indicate that selective filtering of neural response to spurious binaural localization cues is apparent in the DNLL, immediately following the extraction of those cues via coincidence detection in the superior olive.

ACKNOWLEDGMENTS

The authors thank M. Pecka, I. Siveke, and G. Schebesch for assistance in establishing the electrophysiological experiments. A. N. Burkitt, D. B. Grayden, D. R. F. Irvine, N. Lesica, I. Siveke, and D. Taft provided valuable comments on various versions of the manuscript.

Akeroyd, M. A. (**2004**). "The across frequency independence of equalization of interaural time delay in the equalization-cancellation model of binaural unmasking," J. Acoust. Soc. Am. **116**, 1135–1148.

Bauer, B. B. (**1961**). "Phasor analysis of some stereophonic phenomena," J. Acoust. Soc. Am. **33**, 1536–1539.

Bernstein, L. R., Trahiotis, C., Akeroyd, M. A., and Hartung, K. (**2001**). "Sensitivity to brief changes of interaural time and interaural intensity," J. Acoust. Soc. Am. **109**, 1604–1615.

Blauert, J. (**1972**). "On the lag in lateralization cause by interaural time and intensity differences," Audiology **11**, 265–270.

Blauert, J. (**1997**). "Spatial hearing with multiple sound sources and in enclosed spaces," *Spatial Hearing: The Psychophysics of Human Sound Localization*, revised ed. (MIT, Cambridge, MA), pp. 201–287.

Boehnke, S. E., Hall, S. E., and Marquardt, T. (**2002**). "Detection of static and dynamic changes in interaural correlation," J. Acoust. Soc. Am. **112**, 1617–1626.

Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Grothe, B. (**2002**). "Precise inhibition is essential for microsecond interaural time difference coding," Nature (London) **417**, 543–547.

Burger, R. M., and Pollak, G. D. (**2001**). "Reversible inactivation of the dorsal nucleus of the lateral lemniscus reveals its role in the processing of multiple sound sources in the inferior colliculus of bats," J. Neurosci. **21**, 4830–4843.

Caird, D. M., Palmer, A. R., and Rees, A. (**1991**). "Binaural masking level difference effects in single units of the guinea pig inferior colliculus," Hear. Res. **57**, 91–106.

Culling, J. F. (**2008**). "Evidence specifically favoring the equalization-cancellation theory of binaural unmasking," J. Acoust. Soc. Am. **122**, 2803–2813.

Culling, J. F., and Colburn, H. S. (**2000**). "Binaural sluggishness in the perception of tone sequences and speech in noise," J. Acoust. Soc. Am. **107**, 517–527.

Culling, J. F., Hodder, K. I., and Colburn, H. S. (**2003**). "Interaural correlation discrimination with spectrally-remote flanking noise: Constraints for models of binaural unmasking," Acta. Acust. Acust. **89**, 1049–1058.

Culling, J. F., and Summerfield, Q. (**1998**). "Measurements of the binaural temporal window using a detection task," J. Acoust. Soc. Am. **103**, 3540–3553.

Durlach, N. I., and Colburn, H. S. (**1978**). "Binaural phenomena," in *The Handbook of Perception*, edited by E. C. Carterette and M. P. Friedman (Academic, New York).

Eggermont, J. J., Johannesma, P. I. M., and Aertsen, A. M. H. J. (**1983**). "Reverse-correlation methods in auditory research," Q. Rev. Biophys. **16**, 341–414.

Erulkar, S. D. (**1972**). "Comparative aspects of sound lateralization," Physiol. Rev. **52**, 237–360.

Fitzpatrick, D. C., Kuwada, S., Batra, R., and Trahoitis, C. (**1995**). "Neural responses to simple simulated echoes in the auditory brain stem of the unanesthetized rabbit," J. Neurophysiol. **74**, 2469–2486.

Fitzpatrick, D. C., Kuwada, S., Kim, D. O., Parham, K., and Batra, R. (**1999**). "Response of neurons to click-pair as simulated echoes: Auditory nerve to auditory cortex," J. Acoust. Soc. Am. **106**, 3460–3472.

Freyman, R. L., McCall, D. D., and Clifton, R. K. (**1998**). "Intensity discrimination for precedence effect stimuli," J. Acoust. Soc. Am. **103**, 2031–2041.

Grantham, D. W. (**1982**). "Detectability of time-varying interaural correlation in narrow-band noise stimuli," J. Acoust. Soc. Am. **72**, 1178–1184.

Grantham, D. W., and Wightman, F. L. (**1978**). "Detectability of varying interaural temporal differences," J. Acoust. Soc. Am. **63**, 511–523.

Hahn, S. L. (**1996**). "Hilbert transforms," in *The Transforms and Applications Handbook*, edited by A. D. Poularikas (CRC, Boca Raton, FL).

Hirsh, I. J. (**1948**). "The influence of interaural phase on interaural summation and inhibition," J. Acoust. Soc. Am. **20**, 536–544.

Jeffress, L. A. (**1948**). "A place theory of sound localization," J. Comp. Physiol. Psychol. **41**, 35–39.

Jiang, D., McAlpine, D., and Palmer, A. R. (**1997a**). "Responses of neurons in the inferior colliculus to binaural masking level difference stimuli measured by rate-versus-level functions," J. Neurophysiol. **77**, 3085–3106.

Jiang, D., McAlpine, D., and Palmer, A. R. (**1997b**). "Detectability index measures of binaural masking level difference across populations of inferior colliculus neurons," J. Neurosci. **17**, 9331–9339.

Joris, P. X., Schreiner, C. E., and Rees, A. (**2004**). "Neural processing of amplitude-modulated sounds," Physiol. Rev. **84**, 541–577.

Joris, P. X., van de Sande, B., Recio-Spinoso, A., and van der Heijden, M. (**2006**). "Auditory midbrain and nerve responses to sinusoidal variations in interaural correlation," J. Neurosci. **26**, 279–289.

Keller, C. H., and Takahashi, T. T. (**2005**). "Localization and identification of concurrent sounds in the owl's auditory space," J. Neurosci. **25**, 10446–10461.

Klumpp, R., and Eady, H. (**1956**). "Some measurements of interaural time difference thresholds," J. Acoust. Soc. Am. **28**, 215–232.

Kollmeier, B., and Gilkey, R. H. (**1990**). "Binaural forward and backward masking: Evidence for sluggishness in binaural detection," J. Acoust. Soc. Am. **87**, 1709–1719.

Kullback, S. (**1959**). *Information Theory and Statistics* (Wiley, New York).

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.

Litovsky, R. Y., and McAlpine, D. (**2009**). "Physiological correlates of the precedence effect and binaural masking level differences," in *Oxford Handbook of Auditory Sciences*, edited by A. Rees and A. Palmer (Oxford University Press, Oxford), Vol. **2**.

Litovsky, R. Y., and Yin, T. C. T. (**1998a**). "Physiological studies of the precedence effect in the inferior colliculus of the cat. I. Correlates of psychophysics," J. Neurophysiol. **80**, 1285–1301.

Litovsky, R. Y., and Yin, T. C. T. (**1998b**). "Physiological studies of the precedence effect in the inferior colliculus of the cat. II. Neural mechanisms," J. Neurophysiol. **80**, 1302–1316.

Macpherson, E. A., and Middlebrooks, J. C. (**2002**). "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," J. Acoust. Soc. Am. **111**, 2219–2236.

McAlpine, D., Jiang, D., and Palmer, A. R. (**1996**). "Binaural masking level differences in the inferior colliculus of the guinea pig," J. Acoust. Soc. Am. **100**, 490–503.

Maki, K., and Furakawa, S. (**2005**). "Acoustic cues for sound localization by the Mongolian gerbil, meriones unguiculatus," J. Acoust. Soc. Am. **118**, 872–886.

Palmer, A. R., Jiang, D., and McAlpine, D. (**1999**). "Desynchronizing responses to correlated noise: A mechanism for binaural masking level dif-

ferences at the inferior colliculus," J. Neurophysiol. **81**, 722–734.
Pecka, M., Zahn, T. P., Saunier-Rebori, B., Siveke, I., Felmy, F., Wiegrebe, L., Klug, A., Pollak, G. D., and Grothe, B. (**2007**). "Inhibiting the inhibition: A neuronal network for sound localization in reverberant environments," J. Neurosci. **27**, 1782–1790.
Pollak, G. D., Burger, R. M., and Klug, A. (**2003**). "Dissecting the circuitry of the auditory system," TINS **26**, 33–39.
Rayleigh, L. (**1907**). "On our perception of sound direction," Philos. Mag. **13**, 214–232.
Roman, N., Wang, D., and Brown, G. J. (**2003**). "Speech segregation based on sound localization," J. Acoust. Soc. Am. **114**, 2236–2252.
Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (**1967**). "Phase-locked response to low-frequency tones in single auditory nerve fibers of squirrel monkey," J. Neurophysiol. **30**, 769–793.
Shinn-Cunningham, B. G., Santarelli, S., and Kopco, N. (**2000**). "Tori of confusion: Binaural localization cues for sources within reach of a listener," J. Acoust. Soc. Am. **107**, 1627–1636.
Sieveke, I., Pecka, M., Seidl, A. H., Baudoux, S., and Grothe, B. (**2006**). "Binaural response properties of low-frequency neurons in the gerbil dorsal nucleus of the lateral lemniscus," J. Neurophysiol. **96**, 1425–1440.
Siveke, I., Ewert, S. D., Grothe, B., and Wiegrebe, L. (**2007**). "Psychophysical and physiological evidence for fast binaural processing," J. Neurosci. **28**, 2043–2052.
Spitzer, M. W., and Semple, M. N. (**1995**). "Neurons sensitive to interaural phase disparity in gerbil superior olive: Diverse monaural and temporal response properties," J. Neurophysiol. **73**, 1668–1690.
Takahashi, T. T., and Keller, C. H. (**1994**). "Representation of multiple sound sources in the owl's auditory space map," J. Neurosci. **14**, 4780–4793.
Yang, L., and Pollak, G. D. (**1994a**). "Binaural inhibition in the dorsal nucleus of the lateral lemniscus of the mustache bat affects responses for multiplesounds," Aud. Neurosci. **1**, 1–17.
Yang, L., and Pollak, G. D. (**1994b**). "The roles of GABAergic and glycinergic inhibition on binaural processing in the dorsal nucleus of the lateral lemniscus of the mustache bat," J. Neurophysiol. **71**, 1999–2013.
Yin, T. C., and Chan, J. C. (**1990**). "Interaural time sensitivity in medial superior olive of cat," J. Neurophysiol. **64**, 465–488.
Zurek, P. M. (**1987**). "The precedence effect," in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer, New York), pp. 85–105.

Features of across-frequency envelope coherence critical for comodulation masking release

Emily Buss,[a] John H. Grose, and Joseph W. Hall III
Department of Otolaryngology/Head and Neck Surgery, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599

(Received 10 December 2008; revised 7 August 2009; accepted 13 August 2009)

The masking release associated with coherent amplitude modulation of the masker is dependent on the degree of envelope coherence across frequency, with the largest masking release for stimuli with perfectly comodulated envelopes. Experiments described here tested the hypothesis that the effects of reducing envelope coherence depend on the unique envelope features of the on-signal masker as compared to the flanking maskers. Maskers were amplitude-modulated tones (Experiments 1 and 3) or amplitude-modulated bands of noise (Experiment 2), and the signal was a tone; across-frequency masker coherence was manipulated to assess the effects of introducing additional modulation minima in either the on-signal or flanking masker envelopes of otherwise coherently modulated maskers. In all three experiments, the detrimental effect of disrupted modulation coherence was more severe when additional modulation minima were introduced in the flanking as compared to on-signal masker envelopes. This was the case for both ipsilateral and contralateral flanking masker presentations, indicating that within-channel cues were not responsible for this finding. Results are consistent with the interpretation that the cue underlying comodulation masking release is based on dynamic spectral features of the stimulus, with transient spectral peaks at the signal frequency reflecting addition of a signal. [DOI: 10.1121/1.3224708]

PACS number(s): 43.66.Dc, 43.66.Ba [MW] Pages: 2455–2466

I. INTRODUCTION

Detection thresholds for a pure-tone signal in a narrowband masker can be reduced under some stimulus conditions by the introduction of flanking maskers with the same pattern of amplitude modulation as the on-signal masker. This finding, described as comodulation masking release (CMR), has been hypothesized to rely on a wide range of cues, including across-channel processes (Hall *et al.*, 1984; Hall and Grose, 1988; Grose and Hall, 1989; Buus *et al.*, 1996; van de Par and Kohlrausch, 1998b), within-channel cues (Schooneveldt and Moore, 1987; Berg, 1996; Verhey *et al.*, 1999), or some combination of within- and across-channel cues (Schooneveldt and Moore, 1987; Piechowiak *et al.*, 2007). CMR is largest when masker envelopes are perfectly correlated across frequency, but partial correlation supports a reduced CMR effect (McFadden, 1986; Eddins and Wright, 1994). There is evidence that tone detection in coherently modulated maskers is dominated by signal energy coincident with masker modulation minima (Buus, 1985; Hall and Grose, 1991; Buus *et al.*, 1996), consistent with the idea that modulation minima in the flanking maskers are associated with increased stimulus weights during epochs of improved signal-to-noise ratio (SNR); this idea is sometimes described as "cued listening." If the detection advantage conferred by the presence of comodulated flanking maskers is due to off-frequency modulation minima coinciding with epochs of advantageous SNR at the signal frequency, then matches in modulation pattern *per se* may not be critical to CMR, provided that the minima of the off-frequency masker modulation pattern accurately cue minima of the masker at the signal frequency. The present set of experiments tested this hypothesis by manipulating the modulation pattern of on- and off-frequency maskers.

If flanking masker envelope minima cue the optimal listening epochs for signal detection, then introducing additional minima into the pattern of the on-signal masker should have little impact on performance. This manipulation would not disrupt the use of signal information coincident with other on-frequency minima, as cued by the flanking maskers. However, introducing additional minima to off-frequency maskers could provide inappropriate "cues," resulting in increased weight during epochs of disadvantageous SNR and therefore hurting performance. In other words, minima in the on-signal masker that are not present in the flanking maskers should be of little consequence to detection, whereas the converse situation should be very detrimental to performance. One consideration that could impact this prediction is the possibility that a reduction in masker modulation coherence could result in auditory stream segregation (Bregman *et al.*, 1985). Some studies of CMR have proposed that modulation-based auditory grouping may play an important role in CMR (Hall and Grose, 1990; Dau *et al.*, 2004; Dau *et al.*, 2009; Grose *et al.*, 2009), so reduced envelope coherence itself could reduce CMR, quite apart from effects related to the relative timing of specific envelope features. Therefore, while unmatched envelope minima in the on-signal masker may reduce CMR by promoting sound segregation, larger reductions in CMR should be observed with

[a]Author to whom correspondence should be addressed. Electronic mail: ebuss@med.unc.edu

unmatched minima in the flanking maskers, where segregation and inaccurate cueing may combine to reduce sensitivity.

Cued listening is not the only account of CMR consistent with the prediction that the introduction of additional modulation minima to an otherwise coherent pattern of modulation across frequency will have less detrimental effects in the on-signal masker than flanking maskers. In early reports of CMR, the cue underlying detection in a comodulated masker was likened to a dynamic profile analysis (Hall *et al.*, 1984; Fantini and Moore, 1994; Eddins, 2001). In the no-signal interval, the short-term power spectrum of a comodulated masker rises and falls coherently as a function of time, such that the short-term spectrum computed at any point in time is flat as a function of frequency. Addition of a signal disrupts this coherence, in most cases introducing a spectral peak at the signal frequency. For typical CMR stimuli, this cue is most pronounced for signal energy coincident with modulation minima, where the SNR is largest. This cue would also be available for signals that produce a consistent spectral peak and no change in modulation pattern across frequency. For example, Hall and Grose (1988) showed that the detection advantage conferred by coherent modulation does not depend on the signal introducing an envelope discrepancy in the modulation pattern across frequency. In that study, masking release occurred even when the signal was a copy of the narrowband, on-signal masker to which it is added (see also Green and Nguyen, 1988). If a signal in comodulated masking noise is detected by virtue of a spectral peak cue, then a "false positive" could be associated with stimuli in which flanking masker modulation minima are coincident with on-frequency masker modulation maximum. Under these conditions, portions of the on-signal masker could be mistaken for a signal by virtue of the transient spectral peak at the signal frequency that is present even in the absence of an added signal. On the other hand, modulation minima in the on-signal masker envelope that are not associated with flanking masker minima could be of little perceptual consequence. Reductions in level of one tone in the context of a set of otherwise equal-amplitude tones has been shown to be less salient than a level increment (Ellermeier, 1996), suggesting that transient decrements in a dynamic spectral profile could have rather subtle perceptual consequences.

Not all accounts of CMR predict a differential effect of introducing additional minima into the pattern of modulation of the flanking maskers as compared to the on-signal masker. Adding a pure-tone signal to a comodulated masker reduces the coherence of stimulus envelopes across frequency, and it has been suggested that this change could form the basis of the detection cue, quantified as a reduction in envelope correlation or covariance (Richards, 1987; van de Par and Kohlrausch, 1998a). Viewed in this way, CMR could be seen as a special case of monaural envelope correlation perception (Richards, 1987; Strickland *et al.*, 1989). The change in across-frequency envelope coherence with addition of a pure-tone signal in the CMR paradigm has also been characterized as an equalization and cancellation (EC) process (Buus, 1985; Piechowiak *et al.*, 2007), wherein envelopes across frequency are normalized and subtracted, with a residual at the EC output reflecting addition of a signal. Models of the cue underlying CMR, based on envelope decorrelation, and those based on a remainder in an EC process, are fundamentally similar approaches (Green, 1992), both reflecting sensitivity to a reduction in modulation coherence across frequency. A special emphasis on modulation coherence synchronous with modulation minima may be achieved in this modeling approach with compression prior to envelope comparison (Buus *et al.*, 1996). However, there is no basis for a differential effect of additional envelope minima introduced to the flanking as compared to on-signal maskers—in both cases, modulation coherence would be reduced to the same extent by these additional minima.

The experiments reported here measure CMR for two envelope rates, one *slow* and the other *fast*; all minima in slow envelope coincide with minima in the fast envelope, but there are additional minima in the fast envelope that occur during epochs when the slow envelope is high in amplitude. Masker envelopes applied to the on-signal and flanking maskers are either the same (e.g., slow/slow) or different across frequency (fast/slow or slow/fast, defined as the envelope of the on-signal/flanking maskers). Under these conditions, CMR was expected to be largest for conditions in which the masker envelope is the same across frequency. For conditions in which the maskers were not perfectly comodulated across frequency, it was predicted that a fast on-signal masker envelope and slow flanking masker envelope (fast/slow) would be associated with greater CMR than the converse (slow/fast). This expectation is based on the degree to which minima of the flanking masker envelope coincide with modulation minima in the on-signal masker. In the matched envelope conditions, each flanking masker minimum coincides with an on-signal envelope minimum. Similarly, in the fast/slow condition, every minimum in the slow modulation of the flanking maskers coincides with a minimum in the fast on-signal masker modulation. In contrast, in the slow/fast condition, some of the minima in the fast modulation of the flanking maskers coincide with maxima of the slow on-signal masker envelope, resulting in transient spectral peaks at the signal frequency in the short-term spectrum. This could result in false positives because on-signal masker maxima could be confused with an added signal. If all flanking masker modulation minima coincide with on-signal masker minima, as in the matched envelope or fast/slow conditions, the short-term spectrum of the masker would never have a peak at the signal frequency. By this account, the presence of a flanking masker envelope minimum is not always associated with a false positive prediction—just in cases where the on-signal masker is at a high envelope value during a flanking masker minimum.

II. EXPERIMENT 1: EFFECT OF UNMATCHED MINIMA IN ON-SIGNAL VS FLANKING MASKER ENVELOPES FOR TONAL CARRIERS

Stimuli for the first experiment are similar to those used by Grose and Hall (1989). In that study, the signal was a pure tone at 700 Hz, and there were maskers at the 3rd–11th harmonics of 100 Hz, each presented at 50 dB sound pressure

level (SPL). In the coherent modulation condition, all nine masker components were modulated via multiplication with a raised 10-Hz cosine. The signal was composed of three "pips," 50 ms in duration including 20-ms raised-cosine ramps. A masking release was observed if these signal pips were presented synchronously with consecutive envelope minima of the comodulated masker, but thresholds were poor if those pips coincided with masker envelope maxima. Moore *et al.* (1990) performed a follow-up study using stimuli similar to those of Grose and Hall (1989), but in that case omitting the flanking maskers at the sixth and eighth harmonics. This manipulation was designed to assess the contribution of flanking maskers close to the signal frequency, and therefore the effects that might be attributed to within-channel cues. Results were similar with and without those proximal maskers, suggesting that the effects observed with the full complement of maskers are relatively unaffected by within-channel cues. Flanking maskers at the sixth and eighth harmonics were nevertheless omitted in the present study in order to minimize possible effects related to stimulus interaction at the periphery.

A. Methods

1. Observers

Study participants were 13 adults, ages 18–54 (mean of 26.8 years). All had pure-tone detection thresholds of 15 dB hearing level (HL) or less at octave frequencies 250–8000 Hz in the test ear (ANSI, 1996), and none reported a history of ear disease. Eleven observers provided data in the first set of conditions. A subset of five observers went on to complete a second set of conditions, along with two additional observers. These groups were approximately balanced for age and gender. All observers had previously participated in psychoacoustic studies.

2. Stimuli

The signal to be detected was a 700-Hz pure tone, ramped on and off with 25-ms raised-cosine ramps and no steady state. The on-signal masker was an amplitude-modulated (AM) tone at 700 Hz, with the same carrier phase as the signal. Flanking maskers were 100% sinusoidally AM tones at 300, 400, 500, 900, 1000, and 1100 Hz. Both the on-signal and flanking maskers were played continuously over the course of a threshold estimation track, with all carriers starting in sine phase. Two types of envelopes were used, as illustrated in the top two panels of Fig. 1. The fast envelope was a raised 20-Hz cosine. The slow envelope was based on a raised 20-Hz cosine, but the envelope was held at the peak amplitude value for 50 ms on every other period of AM, beginning with the cosine phase of modulation. The signal, when present, was temporally centered in an envelope, minimum of both the fast and slow masker modulation patterns, as illustrated in the bottom panel of Fig. 1.

The flanking maskers were either ipsilateral or contralateral with respect to the ear presented with the signal and the on-signal masker. In the first set of conditions, the peak amplitude of each AM masker tone was 55 dB SPL (the "equal peak" conditions). In the second set, the level of each masker tone with slow AM was reduced by 6 dB, for a peak of 49 dB SPL per masker. This adjustment in level was based on initial data indicating that thresholds in the on-signal masker conditions were approximately 6 dB higher for the slow than the fast AM conditions. As such, this set of conditions will be referred to as the "equal on-signal" conditions.

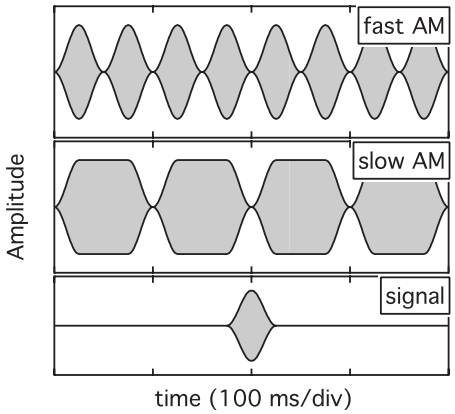


FIG. 1. Basic features of the temporal envelopes used in Experiment 1 are illustrated. The fast pattern of AM was a 20-Hz raised sinusoid, as shown in the top panel. The slow AM stimulus was generated, based on the 20-Hz cosine, with the envelope held constant at the peak amplitude for every other modulation period. The signal was equivalent to a single period of 20-Hz modulation (i.e., 50-ms total duration), with the temporal center of the signal coincident with a minimum in both the fast and slow envelope patterns.

3. Procedures

Stimuli were played out of a real-time processor (RP2, TDT), routed to a headphone buffer (HB7, TDT), and presented over circumaural headphones (Sennheiser HD 265). The experiment was controlled using MATLAB script and RPVDS software (TDT). Observers were seated in a double-walled sound-attenuating booth. A hand-held response box was used to visually indicate listening intervals, collect observer responses, and provide feedback.

Thresholds were measured using a three-alternative forced choice procedure, with each listening interval lasting 350 ms and a 300-ms delay between intervals. When present, the signal coincided with the modulation minimum closest to the temporal center of the listening interval. Feedback was provided visually after every trial. The signal level at the outset of the track was selected to be clearly audible, between 5 and 10 dB above expected threshold. The level was then adjusted based on observer response according to a three-down one-up procedure, estimating the signal level associated with 79% correct (Levitt, 1971). At the beginning of a track, the signal level was adjusted in steps of 4 dB, and steps were reduced to 2 dB after the second track reversal. A track continued for a total of 8 reversals, and a threshold estimate was computed as the average signal level at the last 6 reversals. Three such estimates were obtained in each condition, with a fourth in cases where the first three varied by 3 dB or more, and the final threshold was the average of all such estimates. Observers completed all thresholds in a condition before proceeding to the next condition, and the order of conditions was random across observers.

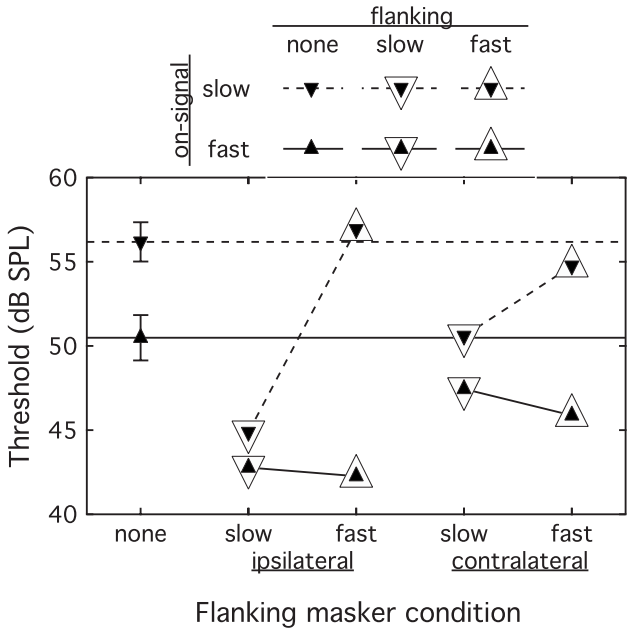


FIG. 2. Mean signal detection thresholds in the equal peak conditions are plotted as a function of flanking masker condition, with filled symbols and line styles reflecting the modulation pattern of the on-signal masker, and open symbols reflecting the modulation pattern of the flanking maskers. Error bars indicate ±1 standard error of the mean.

B. Results and discussion

Mean thresholds across the 11 listeners in the first block of equal peak conditions are plotted in Fig. 2 as a function of flanking masker condition. Symbols and line style indicate the pattern of masker modulation. For the on-signal masker, the slow pattern is indicated by filled down-pointing triangles and dashed lines, and the fast pattern is indicated by filled up-pointing triangles and solid lines. In conditions including flanking maskers, the slow rate is indicated with an open down-pointing triangle, and the fast pattern is indicated by an open up-pointing triangle. The error bars around baseline thresholds show 1 standard error of the mean (sem); no error bars are shown for conditions in which the flanking maskers were present because symbol size exceeds 1 sem.

Thresholds in the absence of flanking maskers (at the left edge of the figure) are about 6 dB lower for the fast than the slow envelope rate. The overall level of the masker in the *fast/none* condition was 2.6 dB lower than that in the *slow/none* condition due to the larger number of modulation mimima. Forward masking could therefore play a role in the difference in on-signal masker thresholds. Inclusion of flanking maskers tended to improve thresholds. As expected, this effect was more pronounced when the masker envelope was comodulated across frequency (slow/slow and fast/fast conditions), and thresholds tended to be lower for ipsilateral than contralateral presentation of the flanking maskers. Thresholds in the fast/slow conditions were relatively low, within several decibels of thresholds in comparable matched-modulation conditions. This was particularly evident in thresholds for ipsilateral flanking masker presentation. In contrast, thresholds were relatively high in the slow/fast conditions. For both ipsilateral and contralateral flanker presentations, mean thresholds in this condition were less than 2 sem away from threshold in the slow/none baseline condition.

TABLE I. CMR, computed as the change in threshold with introduction of flanking maskers relative to the threshold measured with the on-signal masker alone.

		On-signal masker condition	
Flanking masker condition		Fast	Slow
Ipsilateral, equal peak	Fast	8.21	−0.73
	Slow	7.71	11.30
Contralateral, equal peak	Fast	4.61	1.42
	Slow	3.03	5.60
Ipsilateral, equal on-signal	Fast	10.77	−0.54
	Slow	7.35	9.74

The effect of including flanking maskers, described here as the CMR, was computed as the change in threshold relative to the associated no-flanker baseline. Values of CMR, based on the data in Fig. 2, are shown in the top four rows of Table I. The CMR for ipsilateral masker presentation was evaluated using a repeated measures analysis of variance. There were two levels of on-signal masker envelope rate (slow and fast) and two levels of across-frequency masker coherence (same and different). This analysis resulted in a main effect of coherence ($F_{1,10}=87.48, p<0.0001$) and no main effect of on-signal masker envelope rate ($F_{1,10}=2.60, p=0.14$). The interaction was significant ($F_{1,10}=35.11, p<0.0001$). This interaction reflects the fact that a mismatch in modulation patterns across frequency has relatively little effect when the AM associated with the on-signal masker is more rapid than that associated with the flankers (fast/slow) but is pronounced when the modulation pattern associated with the on-signal masker is slower that associated with the flanking maskers (slow/fast).

Estimates of CMR for contralateral masker presentation were also submitted to a repeated measures analysis of variance. As in the previous analysis, there were two levels of on-signal masker envelope rate (slow and fast) and two levels of across-frequency masker coherence (same and different). There was a main effect of coherence ($F_{1,10}=35.77, p<0.0001$) and no main effect of on-signal masker envelope rate ($F_{1,10}=0.03, p=0.87$). The interaction was significant ($F_{1,10}=12.70, p<0.01$), consistent with the conclusion that the interaction observed with ipsilateral flanking maskers was also evident with contralateral maskers.

One aspect of these results that could complicate interpretation of the effects of masker modulation differences across frequency is the masker level discrepancy and the 6-dB difference in baseline, on-signal masking thresholds. To assess the consequences of this difference, a subset of observers repeated the ipsilateral masker conditions of the experiment with the level of all slow masker stimuli reduced by 6 dB, including both slow on-signal and slow flanking maskers. Thresholds for the equal on-signal conditions appear in Fig. 3, and associated values of CMR appear in the bottom two rows of Table I. The CMRs for these conditions

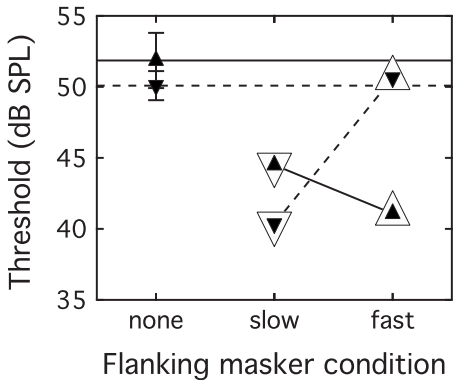


FIG. 3. Mean thresholds for the equal on-signal conditions, with the level of maskers with the slow rate of AM reduced by 6 dB, are plotted as a function of flanking masker condition. Plotting conventions follow those in Fig. 2.

were submitted to a repeated measures analysis of variance, with two levels of on-signal masker envelope rate (slow and fast) and two levels of across-frequency masker coherence (same and different). Results indicate a main effect of on-signal masker envelope rate ($F_{1,10}=8.86, p<0.05$), reflecting better thresholds for the fast envelope rate, and a main effect of coherence ($F_{1,10}=194.63, p<0.0001$), reflecting better thresholds for matched than unmatched rates across frequency. The interaction was also significant ($F_{1,10}=47.47, p<0.01$), confirming that the interaction observed in the original data set was not attributable solely to a discrepancy in baseline masking thresholds for the slow/none and fast/none conditions.

These results are consistent with the initial hypothesis that the effect of introducing envelope pattern discrepancies across frequency depends on the details of that mismatch. In these results, performance appears to be more detrimentally affected by inclusion of additional modulation minima in the flanking than the on-frequency masker envelope. This result is observed when maskers with the two rates are matched for peak amplitude and for maskers approximately matched for threshold in the baseline conditions. The spectral separation of the maskers used in the present study has been argued to reflect across-channel rather than within-channel cues (Moore *et al.*, 1990), providing some support for interpretation of these results in terms of across-channel cues. Whereas similar effects were observed for ipsilateral and contralateral flanking masker presentations, the magnitude of CMR was smaller for contralateral flanking masker presentation. This result could be interpreted as showing that within-channel effects play an important role in the ipsilateral masker results, augmenting a relatively small across-channel CMR.

Whereas some studies of CMR have demonstrated comparable masking release for ipsilateral as compared to contralateral flanking masker presentation (Cohen and Schubert, 1987; Schooneveldt and Moore, 1987), others have reported little or no masking release in the contralateral condition (Hicks and Bacon, 1995; Ernst and Verhey, 2006). It was recently argued that this range of results could be due in part to stimulus features related to auditory grouping (Buss and Hall, 2008). It is widely believed that grouping of the on-signal and flanking maskers into a single auditory stream is a prerequisite for CMR to occur (Dau *et al.*, 2009; Grose *et al.*, 2009). In the present paradigm, it is possible that contralateral masker presentation reduced CMR by reducing auditory grouping of the ipsilateral and contralateral maskers, and further that this effect could have been exacerbated by discrepancies in modulation pattern across maskers in the slow/fast and fast/slow conditions. Therefore, the small contralateral masker effects may not accurately reflect the magnitude of across-channel effects in ipsilateral conditions. This issue will be revisited in the context of Experiment 3.

III. EXPERIMENT 2: EFFECT OF UNMATCHED MINIMA IN ON-SIGNAL VS. FLANKING MASKER ENVELOPES FOR NOISE CARRIERS

The results of Experiment 1 provide support for the idea that the instantaneous level of flanking maskers controls weighting of information at the signal frequency, possibly through cued listening or dynamic spectral profile cues. One limitation of Experiment 1 was the specialized stimulus configuration. Whereas that experiment made use of highly controlled periodic envelopes, this is not representative of more natural auditory stimuli for which masker envelopes are frequently complex. The goal of Experiment 2 was to determine whether the findings of Experiment 1 generalize to a broader set of stimulus conditions.

In Experiment 2 there were two types of signals: a brief signal coincident with a masker modulation minimum and a longer signal that spanned more than one period of modulation. Maskers were bands of noise, either Gaussian noise or low-fluctuation noise, and those bands were multiplied by a raised 10-Hz cosine modulator. These stimuli are analogous to those used in the previous experiment in that the multiplied Gaussian noise contained more numerous prominent modulation minima than the multiplied low-fluctuation noise due to inherent envelope fluctuation, but the long-term power spectra were similar. Both maskers shared envelope minima associated with the 10-Hz periodic modulation. As in Experiment 1, it was hypothesized that additional modulation minima in an otherwise coherent pattern of modulation would be more detrimental when introduced to the pattern of flanking masker AM than when introduced to the on-signal masker AM.

A. Methods

1. Observers

Participants were six adults, ages 19–54 (mean of 31.2 years). All had pure-tone detection thresholds of 15 dB HL or less at octave frequencies 250–8000 Hz in the test ear (ANSI, 1996), and none reported a history of ear disease. All had previously participated in psychoacoustic studies, including one observer who previously participated in Experiment 1.

2. Stimuli

The masker was made up of 30-Hz wide bands of noise, each 60 dB SPL and played continuously throughout a threshold estimation track. All masker bands were multiplied by the same 10-Hz raised cosine. The on-signal band was centered on 1000 Hz, and flanking bands were centered on

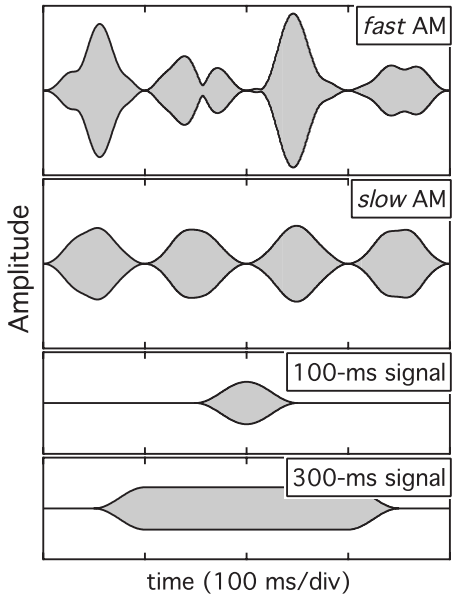


FIG. 4. Basic features of the stimuli used in Experiment 2 are illustrated. The top panel shows a masker with fast amplitude modulation, composed of both relatively rapid inherent fluctuation of a 30-Hz wide band of Gaussian noise and a slower 10-Hz sinusoidal amplitude modulation. The next panel shows a masker with predominantly slow AM and very little inherent modulation, constructed using a sample of low-fluctuation noise. The bottom two panels illustrate the two signal conditions. The brief, 100-ms signal is temporally centered in the minimum of slow masker modulation.

525, 725, 1380 and 1904 Hz. These frequencies were selected based on previous work indicating that within-channel CMR cues related to envelope beats are minimal or absent for log-spaced maskers, and that across-channel cues dominate CMR for a 1000-Hz signal with flanking maskers offset 240 Hz or more from the signal frequency (Grose *et al.*, 2009).

A band of Gaussian noise centered on 1000 Hz was generated in the frequency domain by defining the real and imaginary components within the 30-Hz passband using random draws from a normal distribution. The low-fluctuation noise was generated following procedures described by Kohlrausch *et al.* (1997). Briefly, the Hilbert envelope was extracted from a bandpass noise, and the band was divided by that envelope. The waveform was then transformed into the frequency domain, all components falling outside the original 30-Hz bandwidth were assigned a magnitude of zero, and the result was transformed back into the time domain. This process was repeated ten times. Maskers were generated using arrays that were 2^{17} points long. When played out at a 12207-Hz sampling rate, these arrays played for 10.7 s before repeating seamlessly. Either the low-fluctuation or the Gaussian noise masker at 1000 Hz was loaded into the RPVDS circuit (TDT), where it was AM via multiplication with a raised 10-Hz cosine. The Gaussian masker was loaded in the fast on-signal masker conditions, and the low-fluctuation masker was loaded in the slow on-signal masker conditions. Example stimuli in each condition appear in the top panels of Fig. 4, plotted as a function of time.

Flanking maskers were generated based on the associated 1000-Hz on-signal masker, either low-fluctuation (slow) or Gaussian noise (fast). The values defining the spectral components of the 1000-Hz band were shifted either up or down in frequency and the result transformed to the time domain. The flanking bands were always based on a single set of magnitude and phase values, those associated with either the slow or fast condition. As in the previous experiment, the pattern of modulation was either the same across the on-signal masker and flanking maskers (slow/slow or fast/fast) or different (slow/fast or fast/slow). The array defining the flanking maskers was loaded into the RPVDS circuit and multiplied by the same raised 10-Hz cosine as used in modulation of the on-signal masker. The array defining the flanking maskers was filled with zeros in conditions for which no flanking maskers were present (slow/none and fast/none).

The signal was a 1000-Hz pure tone, gated on and off using 50-ms raised-cosine ramps. In one set of conditions, the offset ramp was initiated as soon as the onset had been completed, such that the signal had a total duration of 100 ms. In a second set of conditions, the offset ramp was initiated 200 ms after completion of the onset ramp, for a total duration of 300 ms. In both cases, initiation of the signal onset ramp coincided with a masker modulation maximum in the slow (10-Hz) pattern of AM, such that the signal reached its full amplitude coincident with the temporal center of a modulation minimum. The relationship between signal presentation and masker modulation is illustrated for each signal condition in bottom two panels of Fig. 4.

3. Procedures

Threshold estimation procedures were identical to those of the previous experiment. Stimuli were played out of a real-time processor (RP2, TDT), routed to a headphone buffer (HB7, TDT), and presented over circumaural headphones (Sennheiser HD 265). Listening intervals were 350 ms, with the signal approximately temporally centered in an interval, and the interstimulus interval was 300 ms. Feedback was provided after every response. Thresholds were collected blocked by condition, but all conditions were run in a different order for each observer.

B. Results and discussion

Figure 5 shows mean thresholds across the six listeners, plotted as a function of flanking masker condition. The left panel shows results for the 100-ms signal, and the right for the 300-ms signal. As in previous data figures, the filled down-pointing triangles and dashed lines indicate a slow on-signal AM, and filled up-pointing triangles and solid lines indicate a fast on-signal AM. Larger unfilled symbols indicate the flanking masker pattern. Error bars indicate 1 sem. No error bars are shown for conditions in which the flanking maskers were present because the symbol size exceeded 1 sem.

For the short duration signal, thresholds are higher for the fast than the slow on-signal masker conditions, with means of 63.3 and 56.2 dB, respectively. This result is in contrast to the results of Experiment 1, where thresholds in the absence of flanking maskers were higher in the slow than

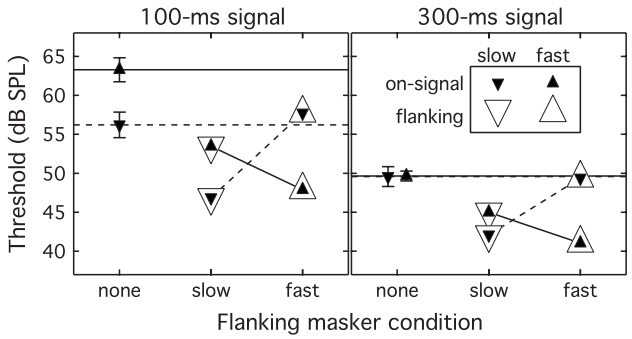


FIG. 5. Mean signal detection thresholds are plotted as a function of flanking masker condition, with symbols and line styles reflecting masker modulation patterns, as defined in the legend. Error bars indicate ±1 sem. The left panel shows results for the brief, 100-ms signal, and the right for the longer, 300-ms signal.

the fast envelope conditions. Thresholds were similar for the two masker types with the 300-ms signal, with mean of 49.6 dB in both the fast and slow modulation conditions. It is not clear why thresholds would differ for the two modulation rates for the brief but not for the longer signal duration. One possibility has to do with the unpredictable envelope events characterizing inherent modulation of the Gaussian noise (fast) masker, events that could be confused for a brief signal. In contrast, for the low-fluctuation (slow) masker, the predictable 10-Hz modulation alone may pose less of a challenge to detection of a brief signal. The long signal may be more easily distinguished from features of the masker in both slow/none and fast/none envelope conditions. The comparable thresholds for the long-duration signal in the two masker types might seem at odds with previous results indicating better threshold in low-fluctuation than Gaussian noise (Hartmann and Pumplin, 1988; Kohlrausch *et al.*, 1997). However, it should be kept in mind that the noises used here were sinusoidally modulated. This additional source of envelope fluctuation could have introduced cues related to low-rate envelope periodicity and/or interfered with cues related to the introduction of envelope fluctuation associated with the addition of a tonal signal to a low-fluctuation noise.

Similar to the results of Experiment 1, thresholds tend to improve with inclusion of flanking maskers in all except for the slow/fast condition, when the envelope of the on-signal masker was slow and the flanking masker envelope was fast. The CMR was computed as the change in thresholds with introduction of flanking maskers, with mean results appearing in Table II. All values of CMR from Experiment 2 were submitted to a repeated measures analysis of variance. There were two levels of on-signal masker envelope rate (fast and slow), two levels of across-frequency masker coherence (same and different), and two levels of signal duration (100 and 300 ms). There was a main effect of coherence ($F_{1,5}=48.35, p<0.001$), a main effect of on-signal masker envelope rate ($F_{1,5}=26.61, p<0.005$), and a main effect of duration ($F_{1,5}=10.05, p<0.05$). All three of the two-way interactions were significant ($6.98 \geq F_{1,5} \leq 14.56, p<0.05$). The three-way interaction was not significant ($F_{1,5}=0.64, p=0.46$). These results replicate the finding of a reduced effect of modulation pattern mismatch when the higher rate modulation is applied to the masker at the signal frequency, and the lack of a three-way interaction demonstrates that this finding is not dependent on signal duration. This can be seen in Fig. 5 as the more consistent detection advantage that is obtained with inclusion of flanking maskers when the on-signal masker has the fast envelope, as compared to the relative variability in masking release when the on-signal masker has a slow envelope.

TABLE II. CMR, computed as the change in threshold with introduction of flanking maskers relative to the threshold measured with the on-signal masker alone.

		On-signal masker condition	
		Fast	Slow
100-ms signal	Fast	15.39	−1.53
	Slow	9.80	9.33
300-ms signal	Fast	8.62	0.19
	Slow	4.66	7.50

The paradigm of Experiment 2 in some ways resembles that of Eddins and Wright (1994). In one set of conditions in that study, pure-tone detection thresholds were measured in a set of five maskers made up of 100-Hz wide bands of Gaussian noise that had been multiplied by a raised 10-Hz cosine. Modulation coherence was defined in terms of the inherent modulation of the bands and/or the phase of sinusoidal AM. Inherent modulation in this experiment could be seen as analogous to the fast modulation in the present experiments, and the 10-Hz sinusoidal modulation as the slow rate. Inherent modulation coherence improved thresholds by approximately 5 dB regardless of the phase of the sinusoidal modulator, and coherence of the sinusoidal modulator improved thresholds by approximately 19 dB regardless of the coherence of inherent modulation. The best thresholds were obtained when both aspects of envelope modulation were coherent across frequency. In contrast to the present experiment, however, both the slow (multiplied) and fast (inherent) modulation patterns were always present in both on-signal and flanking maskers of the Eddins and Wright (1994) study. The present paradigm adds to this work by assessing the detrimental effects of including unmatched modulation minima into the AM pattern of either the on-signal masker or the flanking maskers, rather than both simultaneously.

IV. EXPERIMENT 3: CONDITIONS PROMOTING THE USE OF ACROSS-CHANNEL CUES

Discussion of the results from the first two experiments has focused on the use of across-channel cues. Both cued listening and the use of dynamic spectral cues assume that the beneficial effects of coherent masker modulation are based on a comparison of auditory filter outputs distributed across frequency. The assumption of minimal within-channel effects is justified to some extent by the selection of masker frequencies that have been argued elsewhere to result in primarily across-channel cues (Moore *et al.*, 1990; Grose *et al.*, 2009). However, it is also possible that within-channel cues could have also contributed to masking release. One aspect

of the results that is qualitatively consistent with this possibility is the small effect obtained with contralateral masker presentation in Experiment 1; in those conditions, CMR in the fast/slow conditions was only 1.6 dB larger than CMR in the slow/fast condition (see Table I). This might be interpreted as indicating that the 8.4-dB difference in the CMR observed under comparable ipsilateral conditions was dominated by within-channel effects. However, this result could also reflect reduced auditory grouping associated with contralateral masker presentation (e.g., Buss and Hall, 2008).

The goal of the final experiment was to quantify the CMR obtained with ipsilateral and contralateral flanking maskers under conditions designed to facilitate grouping of the on-signal and flanking maskers. In this paradigm, manipulation of envelope coherence was limited to the listening intervals, and masker modulation was perfectly coherent between trials and between intervals. This stimulus configuration was expected to promote modulation-based grouping based on previous work showing that grouping may be affected by envelope coherence over a relatively protracted period of time (Grose and Hall, 1993; Mendoza *et al.*, 1998; Dau *et al.*, 2009; Grose *et al.*, 2009).

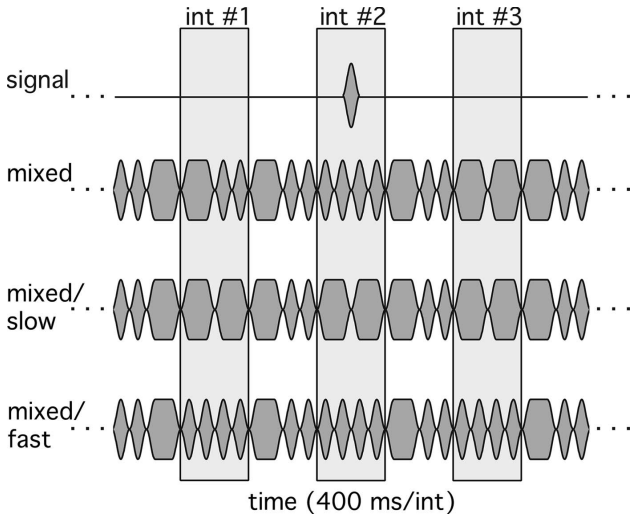


FIG. 6. Basic features of the temporal envelopes used in Experiment 3 are illustrated. The top waveform shows a brief signal presented in the second listening interval. Below that is an example of a mixed masker envelope. The bottom two rows show comparable flanking maskers in the mixed/slow and mixed/fast modulation conditions. Modulation in these conditions is coherent with that of the on-signal masker except inside the three listening intervals, indicated by vertical bars, where the pattern reverted to a consistently slow or fast pattern of AM, respectively.

A. Methods

1. Observers

Participants were six adults, ages 21–54 (mean of 36.9 years). All had pure-tone detection thresholds of 20 dB HL or less at octave frequencies 250–8000 Hz bilaterally (ANSI, 1996), with one exception: one observer had a threshold of 30 dB at 4000 Hz in her left ear. No observer reported a history of ear disease. All had previously participated in psychoacoustical studies unrelated to the present research.

2. Stimuli

The masker was made up of amplitude-modulated tones, each with a peak level of 60 dB SPL and played continuously throughout a threshold estimation track. As in Experiment 2, the on-signal masker was centered on 1000 Hz, and flanking maskers were centered on 525, 725, 1380, and 1904 Hz. Masker modulation in most conditions was unpredictable, comprised of sequences of 100- and 200-ms modulation periods. This *mixed* envelope pattern was generated based on 100-ms time segments. On odd-numbered segments, the envelope was single period of a raised 10-Hz cosine, consisting of a 50-ms offset followed immediately by a 50-ms onset. On even numbered segments, the envelope was either a single period of 10-Hz cosine or a steady "on" plateau; these two possibilities were equally likely, selected based on a random draw from a uniform distribution. The signal, when present, was a copy of the 1000-Hz tone used to generate the on-signal masker, ramped on and off with 50-ms raised-cosine ramps and no steady state. The signal was temporally centered in an odd-numbered envelope segment, ensuring that it coincided with a modulation minimum. The top two waveforms in Fig. 6 illustrate the temporal relationship between the signal and masker modulation. In this example, the signal (top) is presented in the second of three listening intervals, synchronous with a modulation minimum in a series of fast (10-Hz) modulations of the on-signal masker (second trace from the top).

This method of masker modulation intersperses fast and slow modulation periods, while ensuring that the signal is presented under conditions of comparable energetic masking (i.e., an envelope minimum). One consequence of this randomization was to increase thresholds by eliminating masker regularity as a possible detection cue, while holding energetic masking relatively constant. To assess the effects of on-signal masker modulation irregularity, on-signal masker thresholds were also measured in conditions of regular modulation, either consistently fast or consistently slow (not shown in Fig. 6), in addition to the mixed pattern shown in Fig. 6.

Flanking masker conditions were all based on mixed on-signal modulation. In some conditions, the four flanking maskers were presented ipsilateral to the signal and on-signal masker, and in others, they were presented contralaterally. In all cases, the on-signal and flanking maskers were perfectly comodulated between trials and between listening intervals, and conditions differed only with respect to modulation coherence within the three listening intervals of each trial. In the *mixed/mixed* flanking masker condition, the on-signal and flanking maskers were coherently modulated within the listening interval, whereas in the *mixed/fast* and *mixed/slow* conditions, masker modulation coherence was disrupted during the listening intervals. In the mixed/slow condition, the flanking masker modulation during the listening interval was defined with 100-ms plateaus on odd-numbered segments of the masker, regardless of on-signal masker modulation. In the mixed/fast condition, the flanking masker modulation during the listening interval was defined with periods of raised 10-Hz cosine for both odd- and even-numbered masker segments. Examples of the mixed/slow and mixed/

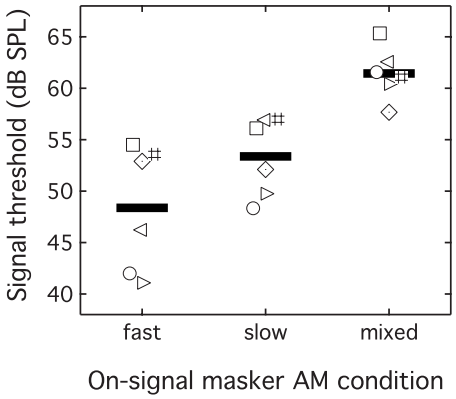


FIG. 7. Mean signal detection thresholds are plotted as a function of the on-signal masker envelope condition. Symbols indicate individual observer's thresholds, and the horizontal bars show the mean threshold.

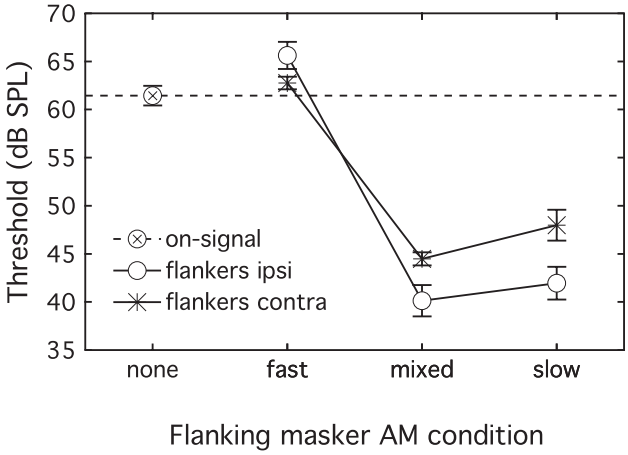


FIG. 8. Mean signal detection thresholds for mixed on-signal AM are plotted as a function of flanking masker condition. The mixed/none on-signal baseline condition is indicated with the filled circle and dashed line. Thresholds in the ipsilateral masker conditions are shown with open circles, and those for the contralateral conditions with asterisks. Error bars indicate ±1 sem.

fast flanking masker conditions are illustrated in the bottom half of Fig. 6.

Both the signal and masker were generated in the RPVDS circuit, including randomization of masker modulation pulse type.

3. Procedures

Threshold estimation procedures were identical to those of the previous two experiments. Listening intervals were 400 ms, separated by 400-ms interstimulus intervals. The signal was exactly temporally centered in one of the three listening intervals, selected at random. Thresholds were collected blocked by condition, but all conditions were run in a different order for each observer.

B. Results and discussion

Results in the on-signal masker conditions will be discussed first. On-signal masker data are shown in Fig. 7, with symbols indicating individual observer's thresholds, and the dark horizontal bars indicating the mean threshold. The ordering of masker conditions along the abscissa was determined according to expected threshold. At the outset, it was expected that the fast/none condition would be associated with lower thresholds than those in the slow/none condition. This expectation was based on the higher masker level associated with the slow rate, and by the results of Experiment 1. It was further anticipated that thresholds in the mixed/none condition would be higher than either of the regular modulation conditions, due to the disruptive effects of stimulus uncertainty (e.g., Watson and Kelly, 1981; Neff and Callaghan, 1988). These expectations were supported by the data, where mean thresholds were 48.4 dB in the fast/none condition, 53.4 dB in the slow/none condition, and 61.4 dB in the mixed/none condition. Most relevant to the present experiment, mixed/none thresholds were worse than those in either of the regular masker modulation conditions. Thresholds were 4.1–13.2 dB worse in the mixed/none than the slow/none condition, a significant result despite the substantial individual differences ($t_5=5.58, p<0.01$). This result cannot be explained in terms of energetic masking and is consistent with the expectation that an irregular pattern of masker modulation eliminates rhythmic cues and introduces non-energetic masking, sometimes described as informational masking (Durlach, 2006).

Figure 8 illustrates the effect of presenting flanking maskers with a mixed on-signal masker. Mean thresholds are plotted as a function of flanking masker modulation condition, indicated on the abscissa. Symbol shape indicates flanking masker presentation condition, which was either ipsilateral (circles), contralateral (asterisks), or none (combined symbol). The dashed horizontal line indicates mean threshold in the mixed/none on-signal masker baseline condition, and error bars indicate 1 sem. The effect of including flanking maskers differed markedly as a function of masker modulation condition. For the mixed/mixed masker modulation conditions, threshold reductions were 21.3 dB for ipsilateral presentation and 17.0 dB for contralateral presentation. This masking release was slightly reduced for the mixed/slow masker modulation conditions, with means of 19.5 and 13.5 dB, respectively. In contrast, flanking maskers with the mixed/fast pattern of modulation failed to reduce thresholds relative to baseline. In these conditions, thresholds rose by 4.2 and 1.3 dB, respectively, though neither of these effects was significantly different from zero as assessed using paired t-tests ($\alpha=0.05$).

These results clearly support the conclusion that the corrupting across-frequency masker modulation coherence depends critically on the details of the across-frequency envelope discrepancy. A masking release on the order of 20 dB is obtained when envelope modulation is perfectly coherent across frequency. There is a relatively small reduction in masking release for conditions in which the on-signal masker modulation includes additional minima that are not present in the pattern of flanking masker modulation. In these conditions, masking release decreased by 1.8 dB (ipsilateral) and 3.5 dB (contralateral). In contrast, introducing additional modulation minima into the pattern of flanking masker modulation is very disruptive, eliminating masking release in the present experiment. Because this effect is seen for both

ipsilateral and contralateral presentation of the flanking maskers, it cannot be attributed to peripheral effects.

While these results are broadly consistent with those of the previous two experiments, the effects are larger, both in terms of the magnitude of masking release and differential effects of introducing additional minima to the on-signal or flanking maskers of an otherwise coherently modulated masker. The large contralateral masking release is particularly noteworthy, in light of the fact that previous studies have reported dichotic CMR of between 0 and 8 dB (Schooneveldt and Moore, 1987; Hicks and Bacon, 1995; Buss and Hall, 2008). One reason for the large masking release observed in the present experiment could be the use of an unpredictable on-signal masker modulation pattern. It has been argued that CMR is largest when performance in the on-signal masker condition is poor due to perceptual similarities between masker fluctuations and fluctuations associated with an added signal (e.g., Moore *et al.*, 1990). This effect could be described in terms of nonenergetic masking. Adding a signal to a masker with predictable modulation, such as periodic slow or fast on-signal modulation, might introduce cues based on a disruption of that regular pattern, whereas this cue would not be available with the mixed on-signal masker modulation. In this case, removing the rhythmic cue could be seen as increasing nonenergetic masking, to the extent that it affects identification rather than encoding of the signal. The effect of introducing masker modulation irregularities in the present experiment is on the order of 10 dB, as reflected in the thresholds measured in the regular (fast/none or slow/none) as compared to irregular (mixed/none) on-signal masker conditions. This could account in part for the approximately 10-dB larger masking release obtained in this experiment, as compared to Experiment 1 and the brief-signal conditions of Experiment 2.

Another aspect of the results that deserves mention is the large effect of contralateral flanking maskers in the present experiment, as compared to Experiment 1. Whereas some of this effect could be due to release from nonenergetic masking related to unpredictability of masker modulation, auditory grouping may also play a role. Previous work has been interpreted as showing that on-signal and flanking maskers must be grouped together in a single auditory stream in order to support CMR, an idea based in part on the observation that CMR can be corrupted by asynchronous onset (Grose and Hall, 1993; Dau *et al.*, 2004) or reduction in envelope coherence outside the listening interval (Grose *et al.*, 2009). Presenting the on-signal and flanking maskers in separate ears could further encourage segregation (Buss and Hall, 2008). In the present paradigm, maskers were coherently modulated across frequency between trials and between intervals even in the mixed/fast and mixed/slow masker AM conditions, a manipulation designed to facilitate grouping of the maskers. The larger masking release obtained in the present experiment could be due in part to stronger auditory grouping, an account which might be particularly applicable to contralateral conditions.

V. GENERAL DISCUSSION

As in previous work, data presented here show that disruption of masker envelope coherence across frequency disrupts CMR. The novel finding of the present study is that the nature of the masker envelope discrepancy across frequency affects the extent to which CMR is disrupted. Inclusion of additional envelope minima into an otherwise coherent pattern of masker modulation across frequency is of less consequence when those minima are introduced to the on-signal masker (fast/slow or mixed/slow conditions) and as compared to the flanking maskers (slow/fast or mixed/fast conditions). In the present conditions, CMR can be eliminated when envelope coherence is corrupted with epochs in which the level of the on-signal masker exceeds that of the flanking maskers, whereas the converse manipulation may only reduce CMR by a factor of 2 or less. The observation that this result is found for contralateral and ipsilateral presentations of the flanking maskers is inconsistent with an interpretation that within-channel processes play a major role in this result. As shown in Experiment 2, this effect is evident for a brief signal, synchronous with a single masker modulation minimum, and for longer signal durations, which span multiple modulation periods. It can also be demonstrated with both spectrally simple and complex patterns of amplitude modulation. The results of Experiment 3 show that the differential effect of disrupting envelope coherence can be quite large, on the order of 20 dB under some conditions, and can be found reliably when flanking maskers are presented contralateral to the signal.

These findings are consistent with previous observations indicating that masker modulation minima are of special significance for CMR with a pure-tone signal, and that use of the associated cues is facilitated by the modulation pattern of the flanking maskers. This effect has been quantified in terms of lower thresholds for brief signals coincident with masker modulation minima (Hall and Grose, 1991) and greater perceptual weight applied to portions of a long-duration signal that coincide with masker modulation minima (Buus *et al.*, 1996). Buus *et al.* (1996) discussed preferential use of information coincident with masker modulation minima in terms of cued listening, whereby modulation minima in flanking maskers indicate "when to listen," but also considered the idea that a similar effect could be obtained by applying a compressive transformation to the stimulus envelope in combination with some other type of across-channel process. After compression, envelope discrepancies occurring during modulation maxima would be small relative to comparable discrepancies during minima. Whereas recent work on within-channel cues available for detection of a tone in coherently modulated maskers have incorporated peripheral nonlinearities, including basilar membrane compression and suppression (Ernst and Verhey, 2008), the role of compression in preferential weighting of modulation minima has not been established. If epochs associated with modulation maxima were compressed relative to minima, this would effectively increase the relative magnitude of signal energy coincident with modulation minima. It would not, however, explain the differential effects of introducing unmatched

modulation minima to the on-signal as compared to the flanking maskers, as observed in the present study.

Data presented here are also qualitatively consistent with the hypothesis that CMR is based on short-term spectral cues, where the cue associated with detection of an added tone can be characterized as a level increment at the signal frequency in an otherwise coherent dynamic spectral profile. In the profile analysis literature, sensitivity to change in a flat spectral profile has been shown to depend on whether that change consists of an increment or decrement. Ellermeier (1996) measured thresholds for detection of either an increment or a decrement in one set of equal-amplitude tones distributed across frequency. He found that decrements were more difficult to detect than increments. This finding could be related to the differential effects of including unmatched modulation minima to on-signal as compared to flanking maskers. When flanking maskers have the slow pattern of modulation, the inclusion additional modulation minima associated with the fast modulation pattern of the on-signal masker could be perceptually subtle, analogous to occasional decrements in relative level at the signal frequency in an otherwise coherent spectrum. In contrast, when the flanking maskers have a fast pattern, the prolonged modulation maxima associated with the slow pattern of the on-signal masker could be perceptually salient, analogous to occasional increments in level at the signal frequency in an otherwise coherent spectrum. Such increments could be mistaken for the addition of a tonal signal, which would also produce occasional spectral peaks at the signal frequency synchronous with modulation minima.

The present results are inconsistent with models of CMR in which the signal is detected based on a change in the pattern of envelope coherence across frequency. Detection cues based on reduced modulation coherence have been quantified as a change in correlation (Richards, 1987; van de Par and Kohlrausch, 1998a) or as a remainder at the output of an EC process (Buus, 1985; Piechowiak *et al.*, 2007). In both cases, introduction of additional modulation minima into an otherwise coherent pattern of masker modulation would have comparable results on envelope coherence, regardless of whether those minima are introduced to the on-signal masker or the flanking maskers. It is possible that envelope comparison models, based on envelope decorrelation or the output of an EC process, could be constructed in such a way as to more heavily weight envelope increments at the signal frequency as compared to decrements. However, it is unclear whether this emphasis would be purely data-driven or whether other theoretical motivations could be identified for such a weighting scheme. The finding of analogous preferential weighting of spectral peaks in the profile analysis literature could lend greater credibility to modeling based on dynamic spectral cues as compared to across-frequency envelope difference cues.

Greater perceptual significance of spectral peaks than dips in the CMR paradigm receives additional support from the results of Moore *et al.* (1990). That study reported thresholds for brief signals presented with a set of sinusoidally modulated masker tones, similar to the stimuli used in Experiment 1. In one set of conditions, the masker tones were 50% AM. The signal was a brief tone, 180° out of phase with the masker tone to which it was added, such that addition of the signal effectively increased modulation depth. Under these conditions, the inclusion of coherently modulated flanking maskers elevated thresholds. Moore *et al.* (1990) argued that this finding is inconsistent with models based on envelope comparison and dip listening, both of which would predict a detection advantage in the presence of comodulated flankers for a signal associated with a decrement in level at the signal frequency. This finding would be more consistent with an explanation of CMR based on a dynamic spectral profile, however, because in this case, the signal would be associated with a relatively subtle spectral dip at the signal frequency.

A special significance for spectral peaks as contrasted with dips has been noted in simultaneous vowel perception (Assmann and Summerfield, 1989) and may represent the preferred mode of processing broadband stimuli in general (Lentz, 2006). While the reason for differential sensitivity to increments and decrements in profile analysis is unknown, Ellermeier (1996) speculated that it could reflect the greater effects of spread of excitation from neighboring tones for a decremented as compared to an incremented target tone. That explanation would not provide a satisfactory account of the present CMR results, however, because the effects of corrupting envelope coherence were evident when flanking maskers were presented contralateral to the signal and on-signal masker. While there are some profile analysis data with contralateral flanker presentation, several authors have argued that these data do not reflect "real" profile analysis because the stimulus components presented to each ear are not fused into a single perceived sound source (Green and Kidd, 1983; Bernstein and Green, 1987). The finding of CMR may also be undermined by segregation effects for dichotic presentation, though continuous presentation and coherent amplitude modulation may counteract these effects to some extent (Buss and Hall, 2008). The role of spectral profile cues in CMR and the differences in sensitivity to spectral peaks and dips is the topic of ongoing research.

VI. SUMMARY

Data reported here support the following conclusions.

(1) Disruption of masker envelope coherence reduces CMR, but the nature of the masker envelope discrepancy across frequency affects the extent of CMR reduction. Masking release is more adversely impacted by inclusion of additional envelope minima to the flanking masker as contrasted with on-signal masker envelopes.
(2) This reduction in masking release is evident for a brief signal, synchronous with a single masker modulation minimum, as well as for longer signal durations, spanning multiple modulation periods. This effect can be demonstrated for both spectrally simple (periodic) and spectrally complex (aperiodic) patterns of amplitude modulation.
(3) The differential effect of including unmatched modulation minima to the on-signal and flanking masker envelopes can be quite large, on the order of 20 dB under

some conditions. These effects are observed for both ipsilateral and contralateral flanking masker presentations, supporting an interpretation in terms of across-channel central processes.

ACKNOWLEDGMENTS

This work was supported by a grant from the NIH NIDCD (Grant No. RO1 DC007391).

ANSI (**1996**). *ANSI S3–1996, American National Standards Specification for Audiometers* (American National Standards Institute, New York).

Assmann, P. F., and Summerfield, Q. (**1989**). "Modeling the perception of concurrent vowels: Vowels with the same fundamental frequency," J. Acoust. Soc. Am. **85**, 327–338.

Berg, B. G. (**1996**). "On the relation between comodulation masking release and temporal modulation transfer functions," J. Acoust. Soc. Am. **100**, 1013–1023.

Bernstein, L. R., and Green, D. M. (**1987**). "The profile-analysis bandwidth," J. Acoust. Soc. Am. **81**, 1888–1895.

Bregman, A. S., Abramson, J., Doehring, P., and Darwin, C. J. (**1985**). "Spectral integration based on common amplitude modulation," Percept. Psychophys. **37**, 483–493.

Buss, E., and Hall, J. W. (**2008**). "Factors contributing to comodulation masking release with dichotic maskers," J. Acoust. Soc. Am. **124**, 1905–1908.

Buus, S. (**1985**). "Release from masking caused by envelope fluctuations," J. Acoust. Soc. Am. **78**, 1958–1965.

Buus, S., Zhang, L., and Florentine, M. (**1996**). "Stimulus-driven, time-varying weights for comodulation masking release," J. Acoust. Soc. Am. **99**, 2288–2297.

Cohen, M. F., and Schubert, E. D. (**1987**). "Influence of place synchrony on detection of a sinusoid," J. Acoust. Soc. Am. **81**, 452–458.

Dau, T., Ewert, S. D., and Oxenham, A. J., (**2004**). "Effects of concurrent and sequential streaming in comodulation masking release," in *Auditory Signal Processing: Physiology, Psychoacoustics and Models*, edited by D. Pressnitzer, A. de Cheveigne, S. McAdams, and L. Collet (Springer-Verlag, Berlin), pp. 335–343.

Dau, T., Ewert, S. D., and Oxenham, A. J. (**2009**). "Auditory stream formation affects comodulation masking release retroactively," J. Acoust. Soc. Am. **125**, 2182–2188.

Durlach, N. (**2006**). "Auditory masking: Need for improved conceptual structure," J. Acoust. Soc. Am. **120**, 1787–1790.

Eddins, D. A. (**2001**). "Monaural masking release in random-phase and low-noise noise," J. Acoust. Soc. Am. **109**, 1538–1549.

Eddins, D. A., and Wright, B. A. (**1994**). "Comodulation masking release for single and multiple rates of envelope fluctuation," J. Acoust. Soc. Am. **96**, 3432–3442.

Ellermeier, W. (**1996**). "Detectability of increments and decrements in spectral profiles," J. Acoust. Soc. Am. **99**, 3119–3125.

Ernst, S. M., and Verhey, J. L. (**2006**). "Role of suppression and retrocochlear processes in comodulation masking release," J. Acoust. Soc. Am. **120**, 3843–3852.

Ernst, S. M., and Verhey, J. L. (**2008**). "Peripheral and central aspects of auditory across-frequency processing," Brain Res. **1220**, 246–255.

Fantini, D. A., and Moore, B. C. J. (**1994**). "A comparison of the effectiveness of across-channel cues available in comodulation masking release and profile analysis tasks," J. Acoust. Soc. Am. **96**, 3451–3462.

Green, D. M. (**1992**). "On the similarity of two theories of comodulation masking release," J. Acoust. Soc. Am. **91**, 1769.

Green, D. M., and Kidd, G., Jr. (**1983**). "Further studies of auditory profile analysis," J. Acoust. Soc. Am. **73**, 1260–1265.

Green, D. M., and Nguyen, Q. T. (**1988**). "Profile analysis: Detecting dynamic spectral changes," Hear. Res. **32**, 147–163.

Grose, J. H., Buss, E., and Hall, J. W. (**2009**). "Within- and across-channel factors in the multiband comodulation masking release paradigm," J. Acoust. Soc. Am. **125**, 282–293.

Grose, J. H., and Hall, J. W. (**1989**). "Comodulation masking release using SAM tonal complex maskers: Effects of modulation depth and signal position," J. Acoust. Soc. Am. **85**, 1276–1284.

Grose, J. H., and Hall, J. W. (**1993**). "Comodulation masking release: Is comodulation sufficient?" J. Acoust. Soc. Am. **93**, 2896–2902.

Hall, J. W., and Grose, J. H. (**1988**). "Comodulation masking release: Evidence for multiple cues," J. Acoust. Soc. Am. **84**, 1669–1675.

Hall, J. W., and Grose, J. H. (**1990**). "Comodulation masking release and auditory grouping," J. Acoust. Soc. Am. **88**, 119–125.

Hall, J. W., and Grose, J. H. (**1991**). "Relative contributions of envelope maxima and minima to comodulation masking release," Q. J. Exp. Psychol. A **43**, 349–372.

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.

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

Hicks, M. L., and Bacon, S. P. (**1995**). "Some factors influencing comodulation masking release and across-channel masking," J. Acoust. Soc. Am. **98**, 2504–2514.

Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., and Oxenham, A. J. (**1997**). "Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations," Acustica **83**, 659–669.

Lentz, J. J. (**2006**). "Spectral-peak selection in spectral-shape discrimination by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 945–956.

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

McFadden, D. (**1986**). "Comodulation masking release: Effects of varying the level, duration, and time delay of the cue band," J. Acoust. Soc. Am. **80**, 1658–1667.

Mendoza, L., Hall, J. W., and Grose, J. H. (**1998**). "Comodulation masking release: The effect of the characteristics of noisebands presented before and after a signal," J. Acoust. Soc. Am. **103**, 2843.

Moore, B. C. J., Glasberg, B. R., and Schooneveldt, G. P. (**1990**). "Across-channel masking and comodulation masking release," J. Acoust. Soc. Am. **87**, 1683–1694.

Neff, D. L., and Callaghan, B. P. (**1988**). "Effective properties of multicomponent simultaneous maskers under conditions of uncertainty," J. Acoust. Soc. Am. **83**, 1833–1838.

Piechowiak, T., Ewert, S. D., and Dau, T. (**2007**). "Modeling comodulation masking release using an equalization-cancellation mechanism," J. Acoust. Soc. Am. **121**, 2111–2126.

Richards, V. M. (**1987**). "Monaural envelope correlation perception," J. Acoust. Soc. Am. **82**, 1621–1630.

Schooneveldt, G. P., and Moore, B. C. (**1987**). "Comodulation masking release (CMR): Effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band," J. Acoust. Soc. Am. **82**, 1944–1956.

Strickland, E. A., Viemeister, N. F., Fantini, D. A., and Garrison, M. A. (**1989**). "Within-versus cross-channel mechanisms in detection of envelope phase disparity," J. Acoust. Soc. Am. **86**, 2160–2166.

van de Par, S., and Kohlrausch, A. (**1998a**). "Analytical expressions for the envelope correlation of narrow-band stimuli used in CMR and BMLD research," J. Acoust. Soc. Am. **103**, 3605–3620.

van de Par, S., and Kohlrausch, A. (**1998b**). "Comparison of monaural (CMR) and binaural (BMLD) masking release," J. Acoust. Soc. Am. **103**, 1573–1579.

Verhey, J. L., Dau, T., and Kollmeier, B. (**1999**). "Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation-filterbank model," J. Acoust. Soc. Am. **106**, 2733–2745.

Watson, C. S., and Kelly, W. J. (**1981**). "The role of stimulus uncertainty in the discrimination of auditory patterns," in *Auditory and Visual Pattern Recognition*, edited by D. J. Getty and J. H. Howard (L. Erlbaum Associates, Hillsdale, NJ), pp. 37–59.

Effects of masker envelope coherence on intensity discrimination

Emily Buss and Joseph W. Hall III
Division of Otolaryngology/Head and Neck Surgery, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599

(Received 18 December 2008; revised 15 June 2009; accepted 4 August 2009)

Masked detection threshold for a pure tone signal depends on the coherence of masker envelope fluctuation across frequency, with lower thresholds for coherent fluctuation under some conditions. The benefit of coherent masker modulation is larger for detection than for suprathreshold tasks, such as pure tone intensity discrimination [Hall, J. W. and Grose, J. H. (1995). J. Acoust. Soc. Am. **98**, 847–852]. In the present study, sensitivity to increments in signal intensity was measured for a 1000-Hz signal, either a tone or a 20-Hz-wide narrowband noise. In one set of conditions the masker was one or more bands of noise, each 20 Hz wide, and in another set of conditions the masker was a single 1620-Hz-wide band of Gaussian noise or noise multiplied by the envelope of a 20-Hz bandpass noise. Coherent masker envelope fluctuation improved detection thresholds in all conditions. Intensity discrimination for a tonal standard in comodulated noise was elevated for standard levels near detection threshold and improved with increasing signal-to-noise ratio, whereas performance was uniformly poor across level for the noise standard. Results are most consistent with the interpretation that the reduced benefit of coherent masker modulation in suprathreshold intensity discrimination is due to the disruptive effects of envelope fluctuation.
 [DOI: 10.1121/1.3212944]

PACS number(s): 43.66.Dc, 43.66.Fe [MW] Pages: 2467–2478

I. INTRODUCTION

Detection threshold for a tone in narrowband noise can be reduced with the introduction of additional off-frequency noise bands, provided those bands have the same pattern of envelope fluctuation as the on-signal masker band (Hall *et al.*, 1984). This result is sometimes described as comodulation masking release (CMR). Whereas detection thresholds improve with the introduction of coherently modulated flanking maskers, several lines of evidence indicate that masking release is smaller for suprathreshold discrimination than for detection tasks. For example, Hall and Grose (1995) reported greater masking release for detection than for intensity discrimination of a tonal signal presented at a low sensation level, a task described as suprathreshold intensity discrimination.[1] This difference in masking release was illustrated by comparing intensity discrimination thresholds for standard levels that were defined relative to detection threshold. When performance is compared across masker conditions for signals at a low sensation level (SL), such as 10 dB SL, discrimination thresholds are worse in the context of coherently modulated masker bands than in the context of one or more random noise maskers. The CMR obtained for detection is larger than that for suprathreshold discrimination across a range of tasks, including intensity discrimination (Hall and Grose, 1995), gap detection (Hall and Grose, 1992), pitch ranking (Hall *et al.*, 1997), and speech perception (Grose and Hall, 1992). Hall *et al.* (1997) hypothesized that coherent masker envelope fluctuation lowers detection threshold, but that the resulting representation of the signal is degraded, which in turn limits sensitivity in suprathreshold tasks. The purpose of the present experiments was to more fully characterize the differential effects of coherent masker modulation on suprathreshold intensity discrimination as compared to signal detection.

In some cases threshold improvement associated with masker fluctuation coherence is attributed to an across-channel process, whereby detection of the signal is enhanced by information carried in independent auditory channels (Haggard *et al.*, 1990; Hall *et al.*, 1993; Moore *et al.*, 1993; Buss and Hall, 2008; Grose *et al.*, 2009). This process may involve a comparison of stimulus envelopes across frequency or a listening strategy whereby off-frequency modulation is used to facilitate "listening in the dips" of the masker (Buus, 1985). Hall and Grose (1995) suggested that derived detection cues, such as those based on the introduction of across-channel envelope differences, might by their very nature contain less detailed signal information than cues based solely on within-channel information. This suggestion is broadly consistent with the finding of relatively poor suprathreshold intensity discrimination as compared to signal detection in both the CMR and binaural masking level difference paradigms (Henning, 1991; Hall and Grose, 1995). As in the case of monaural masking release, binaural masking release is thought to be based on a comparison of stimuli falling in independent peripheral channels (in this case different ears). Hall and Grose (1995) proposed that such difference cues might result in an internal representation of the signal that is essentially different from that associated with baseline conditions, and they described this degraded representation as "coarse."

Whereas across-channel cues are thought to underlie the masking release in many CMR paradigms, under other conditions within-channel cues may support improved detection threshold in the presence of coherently modulated maskers.

For example, in many cases the within-channel pattern of envelope beats associated with a pair of comodulated narrow bands of noise is disrupted by the addition of a pure tone signal (Schooneveldt and Moore, 1987; Berg, 1996). These within-channel cues could be more useful in detection than discrimination. Further, the proximity of flanking masker bands could interfere with the use of cues that would otherwise contribute to intensity discrimination. In general, intensity discrimination for a spectrally narrow signal tends to improve as a function of level of the standard, a result that has been described as the "near miss" to Weber's law. It is widely accepted that this near miss is due to spread of excitation at higher stimulus levels and hence a wider range of frequency channels representing the signal (e.g., Florentine and Buus, 1981). To the extent that internal noise is independent across channels, this effect could be due to reduced effects of internal noise with multiple samples. Nonlinear growth of masking on the high frequency side of the excitation pattern may also contribute to the beneficial effects of spread of excitation (Zwicker, 1970). Whereas spread of excitation could improve intensity discrimination, masking of this spread could play a role in the relatively poor suprathreshold performance observed with coherently modulated flanking maskers. While random flanking maskers could interfere with intensity discrimination, this effect might be more pronounced in coherent masker conditions where dynamic spectral masking is synchronous above and below the signal frequency.

Variability in the stimulus level associated with masker envelope fluctuation could also play a role in suprathresold intensity discrimination. A classic study by Bos and de Boer (1966) showed that the inherent amplitude modulation (AM) of narrowband noise stimuli limits intensity discrimination as well as sensitivity to an added tone for bandwidths of 10–40 Hz. Corroborating this result, Eddins (2001) reported lower thresholds for intensity increment detection when the standard stimulus was a low-fluctuation noise as compared to a Gaussian noise of the same narrow bandwidth, a result that confirms the effects of stimulus fluctuation independent of bandwidth. When a pure tone signal is added to a bandpass Gaussian noise it tends to flatten the envelope, with greater effects at higher signal-to-noise ratios (SNRs). Assuming that the signal and the masker band centered on the signal are spectrally resolved from flanking maskers, increasing the signal level has a uniform effect on the statistical properties of the encoded signal irrespective of the presence of flanking masker bands. To the extent that coherence of the masker envelope across frequency reduces detection thresholds, stimuli presented at a low SL (e.g., 10-dB SL) will have more pronounced envelope fluctuation in the coherent modulation as compared to the baseline conditions of the CMR paradigm. The detrimental effects of stimulus envelope fluctuation on intensity discrimination could therefore be responsible for the poor suprathreshold discrimination thresholds in coherent masker conditions.

One way to think about the disruptive effects of envelope fluctuation on intensity discrimination is in terms of the differential effects of internal and external noise (Swets *et al.*, 1959). Whereas intensity discrimination thresholds for stationary stimuli are likely to be limited by internal noise, stimuli with an interval-by-interval level rove are limited by external, stimulus-based noise (Spiegel *et al.*, 1981; Jesteadt *et al.*, 2003). In a recent study it was argued that the poorer intensity discrimination for stimuli with fluctuating envelopes might likewise be attributable to increased external noise (Buss *et al.*, 2006). This could have implications for the utility of cues present in the auditory channel centered on the signal as well as the ability to benefit from spread of excitation. If the spread of excitation with increasing presentation level improves intensity discrimination due to a reduction in the effects of internal noise, then these beneficial effects could be more pronounced under conditions of minimal envelope fluctuation, for which internal noise limits performance, as compared to highly fluctuating stimuli, for which external noise limits performance.

Masking noise does not have to be presented synchronously with the signal to interfere with intensity discrimination. Several groups have demonstrated that intensity discrimination can be disrupted by nonsimultaneous masking noise even if that noise does not affect detection threshold (e.g., Zeng *et al.*, 1991; Carlyon and Beveridge, 1993; Plack *et al.*, 1995; Plack, 1996; Oberfeld, 2008). Intensity discrimination in nonsimultaneous masking conditions appears to be quite complex and is not fully understood (see Zeng, 1998; Oberfeld, 2008), but one possible explanation is that maskers presented before or after the signal could interfere with the trace memory representing the intensity in each interval, particularly if the masker has prominent inherent fluctuation (Plack *et al.*, 1995). In the case of intensity discrimination in the presence of coherently modulating masker bands, variability in masker level could interfere with intensity discrimination for standard tones presented near threshold by virtue of corrupting the memory trace associated with the standard. In baseline conditions, with one or more random noise maskers, this interference may be less severe because the level of the standard is closer to the peak masker level, which may serve as a perceptual reference point or anchor (Braida *et al.*, 1984). Disruption of memory traces could also be related to the effects of modulation within the listening interval, with greater disruption under conditions for which the masker and signal-plus-masker stimuli are both highly modulated and therefore perceptually similar.

The broad goal of the experiments described in the present report was to differentiate among possible explanations for reduced masking release in suprathreshold intensity discrimination tasks. To summarize, those included (1) coarse representation of the signal due to the loss of information inherent in an across-channel comparison, (2) masking associated with flanking masker bands, (3) detrimental within-channel effects of stimulus envelope fluctuation, (4) limits on the ability to benefit from spread of excitation due to envelope fluctuation, and (5) detrimental effects of level fluctuation between intervals on memory for the standard level. A secondary goal was to replicate intensity discrimination results previously demonstrated with narrowband maskers and to determine whether similar effects are obtained with a single spectrally contiguous masker, wherein coherent AM is introduced via multiplication by a low-rate

modulator. This spectrally contiguous, multiplied masker was of interest because it more closely resembles comodulated background noise that may be encountered under natural listening conditions.

Experiment 1 measured detection and suprathreshold intensity discrimination thresholds for a pure tone signal presented in either a set of narrowband maskers or a single bandpass masker to determine the effects of coherent AM. These results establish the basic effect under study. Using just narrowband noise maskers, Experiment 2 measured detection and intensity discrimination thresholds for a narrowband noise signal. In contrast to the envelope modulation reduction as a function of SNR obtained with a tonal signal, there is pronounced envelope fluctuation irrespective of SNR when the signal is a narrowband noise. If envelope fluctuation is responsible for the level effects with a tonal signal, then discrimination with narrowband noise signals should be insensitive to level. Experiment 3 explored the relative contributions of envelope variability and spread of excitation to suprathreshold intensity discrimination measured with and without narrowband flanking masker bands.

II. GENERAL PROCEDURES

A. Observers

Observers were 15 normal hearing adults, ages 18–53 years (mean 30 years). All were screened for normal hearing in the test ear, defined as thresholds of 15-dB hearing level (HL) or better for pure tones 250–8000 Hz (ANSI, 1996). None of these observers reported a history of ear disease, and all had previously participated in psychoacoustic studies. A subset of observers completed each experiment, as indicated below.

B. Stimulus generation

The signal was either a 1000-Hz pure tone (Exp 1 and 3) or a 20-Hz-wide band of noise arithmetically centered on 1000 Hz (Exp 2). This signal was gated on and off with 50-ms raised-cosine ramps and had a total duration of 450 ms. In some cases the task was to detect the presence of a signal, while in others the task was to select the interval associated with the most intense signal. Maskers were one or more 20-Hz-wide bands of noise, a single bandpass noise, or an amplitude modulated bandpass noise.

Narrowband and bandpass Gaussian noise samples were generated in the frequency domain, with draws from a normal distribution defining the real and imaginary components within the passband. Coherently modulated narrowband maskers were generated with a single family of random draws to define corresponding components of each band, whereas random bands were generated with independent random draws. Stimuli were then transformed into the time domain with an inverse fast Fourier transform. Each masker array was composed of 2^{17} points which, when played out at 12 207 Hz, could be repeated seamlessly with one repetition every 10.7 s. The average masker presentation levels are reported separately for each experiment. Due to the random fluctuations of noise, the masker level in each listening interval deviated slightly from that mean. Maskers played continuously in Experiments 1 and 2, whereas in Experiment 3 some portions of the masker were gated on only during the listening interval. The maskers were generated in MATLAB prior to every threshold estimation run.

Stimuli were played out at 12 207 Hz (RP2, TDT), passed through a headphone buffer (HB7, TDT), and presented to the left channel of a pair of circumaural headphones (Sennheiser, HD 265).

C. Procedures

All thresholds were estimated using a three-alternative forced-choice procedure and a three-down one-up tracking rule estimating 79% correct (Levitt, 1971). In all cases the masker was held at a constant level over the course of a track, and the level of the signal was adjusted. For the detection task signal level was defined in units of dB SPL, and in the intensity discrimination task the signal was defined in units of $10 \log(\Delta I/I)$. For both detection and intensity discrimination tasks the initial signal level adjustments were made in steps of 4 dB, and steps were reduced to 2 dB after the second track reversal. A total of eight reversals was obtained in each track, and the threshold estimate was the average signal level at the last six track reversals.

In the detection task the observer was presented with three listening intervals, each 450 ms in duration and separated by 300-ms interstimulus intervals. Each listening interval was visually indicated with a light mounted above the associated response button on a handheld response box. The signal was presented in one of these intervals with equal probability, and the observer indicated which interval contained the signal; visual feedback was then provided. In the discrimination tasks the procedures were identical except that there was a standard stimulus in all three intervals, and the observer's task was to select the interval in which the level of that standard was incremented. For intensity discrimination in Experiments 1 and 2, the level of the standard was set relative to each observer's detection threshold in each condition: standard levels were either 10, 20, or 30 dB SL. Standard levels for Experiment 3 were uniform across observers, spaced at 10 dB increments between 50 and 80 dB SPL. Detection thresholds were measured prior to discrimination thresholds in Exp 1 and 2; aside from that constraint, conditions were completed in random order within an experiment.

III. EXPERIMENT 1

The first experiment assessed intensity discrimination for a pure tone in the presence of a comodulated masker as compared to baseline conditions, with one or more bands of independent noise. In one set of conditions the masker was composed of up to five narrow bands of noise. These conditions closely resemble those of Hall and Grose (1995), where suprathreshold intensity discrimination was shown to be poorer in coherent masker conditions than in baseline conditions when comparing performance for a standard presented at a fixed level relative to detection threshold. Other conditions in the present experiment measured intensity discrimination thresholds with a single contiguous bandpass masker

TABLE I. Mean detection thresholds for each masker type examined in Experiment 1, with standard error of the mean shown in parentheses.

	Narrowband			Bandpass	
	On-signal	Random	Coherent	Gaussian	AM-noise
Primary conditions	51.8 (0.34)	51.6 (0.39)	43.0(0.98)	52.1 (0.50)	45.8(0.74)
+10-dB masker level			52.5(1.46)		55.8(0.59)

that was either Gaussian noise or AM noise. It was hypothesized that the effect of masker coherence on suprathreshold intensity discrimination for the bandpass masker would be similar to that previously shown in the narrowband masker paradigm. Such a result would lend support to the idea that the finding of poor suprathreshold discrimination in coherently modulated narrowband maskers may generalize to more natural listening conditions, such as speech masked by a spectrally contiguous fluctuating background noise.

A. Observers

Observers 1–7 participated, including four males, and the mean age in this subgroup of observers was 36 years.

B. Stimuli

The signal was a 1000-Hz pure tone, and the task was to detect the presence of a signal or an increment in the level of the signal. There were five primary masker conditions, three for which the masker was comprised of narrow bands of noise and two in which the masker was a single, spectrally contiguous bandpass noise.

In the *on-signal* masker condition there was a single 20-Hz-wide band of Gaussian noise centered on 1000 Hz. In the *random* masker condition there was a set of five 20-Hz-wide bands of Gaussian noise, centered on 200, 600, 1000, 1400, and 1800 Hz. The *coherent* masker condition included 20-Hz-wide bands at the same frequencies, but those bands were comodulated. In the primary conditions each masker band was presented at 50 dB SPL for an overall level of 57 dB SPL when all five bands were present.

There were two bandpass masker conditions, wherein the masker was filtered to span the same spectral range as maskers in the narrowband noise conditions (190–1810 Hz). In the *Gaussian* condition the masker was a band-limited Gaussian noise. In the *AM-noise* condition a bandpass Gaussian noise sample was multiplied by the Hilbert envelope associated with a 20-Hz narrowband noise, generated using procedures described above for the *on-signal* masker. Bandpass maskers were played at 65 dB SPL.

The masker levels used in the primary conditions described up to this point were chosen to produce approximately equal thresholds in the *on-signal*, *random*, and *Gaussian* baseline conditions. Thresholds were expected to be significantly lower in the *coherent* and *AM-noise* conditions, a reduction associated with introduction of coherent masker envelope fluctuation across masker frequency. In order to allow comparison of intensity discrimination across conditions at an approximately matched signal level, additional data were collected with a 10-dB higher masker level in the two conditions associated with masking release. The *coherent*+10 condition was identical to the *coherent* condition described above, but the masker was presented at an overall level of 67 dB SPL. Similarly, the *AM-noise*+10 condition was identical to the *AM-noise* condition in all respects other than the 75 dB SPL overall masker presentation level. These levels were chosen based on pilot data indicating masking release on the order of 10 dB in both the narrowband and bandpass noise conditions.

C. Results

The pattern of results was broadly consistent across observers, so only mean results will be presented. The mean detection thresholds are reported for each masker condition in Table I. For narrowband maskers, thresholds in the two baseline conditions were quite similar, with means of 51.8 and 51.6 dB in the *on-signal* and *random* conditions, respectively. Thresholds dropped to 43.0 dB in the *coherent* condition for a masking release of approximately 8.7 dB. For bandpass maskers, thresholds in the *Gaussian* baseline condition were 52.1 dB as compared to 45.8 dB in *AM-noise* condition for a masking release of 6.3 dB. Increasing the masker level by 10 dB elevated thresholds by 9.5 dB in the *coherent* condition and by 10.0 dB in the *AM-noise* condition.

Figure 1 shows mean intensity discrimination thresholds plotted in units of 10 $\log(\Delta I/I)$ as a function of the level of the standard tone relative to detection threshold. Masker conditions are indicated with symbols, as shown above each panel. Results for the primary narrowband noise conditions appear in the far left panel (A). In the *on-signal* and *random*

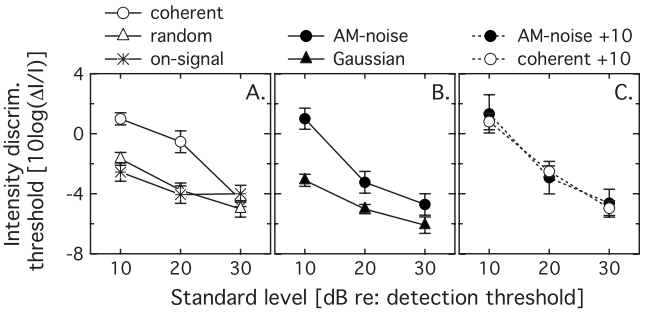


FIG. 1. Mean intensity discrimination thresholds for a pure tone signal are plotted as a function of standard level relative to detection threshold. Error bars indicate ±1 standard error of the mean across the seven observers. Symbols reflect the masker condition, as indicated above each panel. The left panel (A) shows thresholds in the narrowband masker conditions. The middle panel (B) shows bandpass masker results. The far right panel (C) shows thresholds in the two conditions with comodulated maskers played at an increased masker level to approximately counteract masking release.

conditions, thresholds improve by an average of 2.4 dB between 10 and 30 dB SL. In contrast, thresholds in the *coherent* condition improve by 5.4 dB over the same range, converging with the other two conditions at the 30-dB SL standard level. A similar pattern is seen in the primary bandpass masker conditions shown in the middle panel (B). Thresholds in the *Gaussian* noise masker improved by approximately 3.0 dB between 10 and 30 dB SL, whereas those in the *AM-noise* improved by 5.7 dB over that range. Thresholds for the coherent masker conditions with the elevated masker level are shown in the far right panel (C). As in the primary conditions, thresholds in both the *coherent*+10 and *AM-noise*+10 conditions were elevated relative to the no-release conditions at the lowest standard level of 10 dB SL (as illustrated in panels A and B) and converged toward no-release performance with increasing standard level; thresholds in these conditions improved by an average of 5.9 dB between 10 and 30 dB SL. These results indicate that absolute signal level was not responsible for the masker effects noted in the primary data.

Intensity discrimination thresholds in the narrowband noise masker conditions were submitted to a repeated-measures analysis of variance (ANOVA) with three levels of MASKER (*coherent*, *random*, and *on-signal*) and three levels of STANDARD (10-, 20-, and 30-dB SL). There was a main effect of MASKER ($F_{2,12}=33.37, p<0.0001$) and of STANDARD ($F_{2,12}=46.30, p<0.0001$), as well as a significant interaction ($F_{4,24}=6.88, p<0.001$). Preplanned contrasts indicate that the MASKER-by-STANDARD interaction was significant comparing just the *coherent* and *random* conditions ($p<0.01$), but not when comparing the *random* and *on-signal* conditions ($p=0.06$). These results confirm the elevation of intensity discrimination thresholds in the *coherent* condition at the lowest standard level, with reduced masker effects at higher signal levels.

A similar analysis was performed with intensity discrimination thresholds in the bandpass masker conditions, with two levels of MASKER (*AM-noise* and *Gaussian*) and three levels of STANDARD (10-, 20-, and 30-dB SL). There was a main effect of MASKER ($F_{1,6}=25.13, p<0.005$) and of STANDARD ($F_{2,12}=54.44, p<0.0001$) as well as a significant interaction ($F_{2,12}=9.68, p<0.005$). These results confirm that the improvement in intensity discrimination with increasing standard level in dB SL is not uniform across *Gaussian* and *AM-noise* masker types, a result which parallels the narrowband masker results.

The effect of lower absolute level of the standard tone in the coherent masker conditions was assessed with a third analysis. This repeated-measures ANOVA included two levels of MASKER (*coherent* and *AM-noise*), two levels of LEVEL (primary and +10 dB), and three levels of STANDARD (10-, 20-, and 30-dB SL). There was a main effect of STANDARD ($F_{2,12}=95.50, p<0.0001$), but no effect of MASKER ($F_{1,6}=0.33, p=0.59$) or LEVEL ($F_{1,6}=1.52, p=0.26$). There was a significant interaction between STANDARD and MASKER ($F_{2,12}=5.05, p<0.05$), reflecting the trend for thresholds to be 0.4 dB lower in the bandpass than narrowband masker conditions. No other interactions were significant ($\alpha=0.05$). These results confirm the visual impression that increasing masker level, and therefore signal threshold, does not substantially change the pattern of intensity discrimination as a function of standard level relative to detection threshold for these stimuli.

D. Discussion

The coherent masker results of Experiment 1 replicate the general findings of Hall and Grose (1995), where suprathreshold intensity discrimination was shown to be poorer in comodulated narrowband noise maskers than in baseline conditions when compared as a function of signal level relative to detection threshold (dB SL). This result generalized to a single contiguous bandpass noise masker, with better suprathreshold intensity discrimination for a low-SL tone in a Gaussian noise masker than in a noise that was amplitude modulated via multiplication with the envelope of a 20-Hz-wide bandpass noise. The pattern of intensity discrimination thresholds was insensitive to a 10-dB increase in masker level, indicating that suprathreshold performance in the primary conditions was not due to the lower absolute level of the standard tones in masking release conditions. Plotting thresholds as a function of level in dB SL highlights the difference between coherent modulation and baseline conditions. As also noted by Hall and Grose (1995), this difference would be deemphasized by plotting thresholds as a function of SNR. This observation on the importance of units in the comparison of intensity discrimination thresholds across masker conditions is revisited in discussion of the third experiment.

IV. EXPERIMENT 2

Intensity discrimination is poorer for a stimulus that fluctuates randomly in amplitude than for one with a more steady envelope (Bos and de Boer, 1966; Eddins, 2001). When a tone is added to a band of noise, the envelope of the summed stimulus becomes flatter with increasing intensity of the tone. In Experiment 1, intensity discrimination was compared across masker conditions at a low sensation level. Because detection thresholds were higher in baseline than in coherent masker conditions, this resulted in higher standard levels and more envelope flattening in baseline conditions. Therefore, lower signal levels and greater envelope fluctuation in the coherent masker conditions of Experiment 1 could play a role in the relatively poor intensity discrimination and reduced suprathreshold masking release.

Experiment 2 used narrowband noise maskers and examined the role of stimulus fluctuation in suprathreshold intensity discrimination by measuring intensity discrimination for a narrowband noise signal. Because this signal is itself associated with inherent amplitude modulation, the envelope of the masker-plus-signal does not become flatter at increasing SNRs. If level variability of the summed stimulus limits intensity discrimination with a pure tone signal, then signal level should have little or no effect on performance for the narrowband noise signal, and intensity discrimination should be comparably poor across baseline and masking release conditions.

TABLE II. Mean detection thresholds for each masker type examined in Experiment 2, with standard error of the mean shown in parentheses.

On-signal	Random	Coherent-ran	Coherent-copy
54.0 (0.75)	58.3 (1.05)	40.9 (0.99)	49.1 (0.90)

A. Observers

Six observers participated in this experiment, including Obs 6 and Obs 8–12. There were two males in this group, and the mean age was 24 years.

B. Stimuli

The signal was a 20-Hz-wide band of Gaussian noise arithmetically centered on 1000 Hz. There were three masker conditions, identical to the *on-signal*, *coherent*, and *random* narrowband noise conditions described above for Experiment 1. Each masker band was presented at 50 dB SPL for an overall level of 57 dB SPL when all five bands were present. In the primary conditions the signal band was a copy of the masker centered on 1000 Hz, to which it was added. In an additional condition the signal was a random band, independent of the masker band to which it was added, and all maskers were coherently modulated; this condition will be referred to as *coherent-ran* to distinguish it from the *coherent-copy* condition, where the signal was a copy of the masker band. Whereas the pattern of modulation across frequency is unchanged with addition of the signal in the *coherent-copy* condition, envelope coherence is reduced by addition of a signal in the *coherent-ran* condition. These conditions allow an assessment of the importance of across-frequency envelope coherence for suprathreshold intensity discrimination.

C. Results

Detection thresholds for the narrowband noise signal are reported in Table II. In contrast to the results of Experiment 1, thresholds were lower in the *on-signal* than in the *random* condition, with means of 54.0 and 58.3 dB, respectively. This 4.3-dB difference was significant ($t_5=5.38, p<0.005$). Thresholds in these conditions exceeded those measured under analogous conditions with a pure tone signal in Experiment 1 (as reported in Table I) for both the *on-signal* (2.2 dB; $t_{11}=2.83$, $p<0.05$) and the *random* (6.7 dB; $t_{11}=6.43$, $p<0.0001$) conditions. Thresholds improved to 40.9 dB with inclusion of coherently modulated flanking masker bands in the *coherent-ran* condition, comparable to the 43.0-dB threshold for a pure tone in Experiment 1 ($t_{11}=1.52, p=0.16$). Masking release in the *coherent-ran* condition was 13–17 dB, depending on choice of baseline. Sensitivity was not as good in the *coherent-copy* condition, where the mean threshold was 49.1 dB, and the corresponding masking release was 5–9 dB.

Figure 2 shows mean intensity discrimination thresholds plotted in units of $10\log(\Delta I/I)$ as a function of the level of the standard relative to the corresponding detection threshold. Following the conventions of Fig. 1(A), masker conditions are indicated with symbols. Notice that whereas the ordinate scale has been maintained across Figs. 1 and 2 (2 dB/div), the range of ordinate values is shifted up by 4 dB in Fig. 2. The bulls-eye symbols representing thresholds in the *coherent-copy* condition have been offset to the left to improve visual resolution of the data pattern. The most notable aspect of these results is the relative stability of thresholds as a function of standard level and the similarity of results across masking conditions. All means fall in the range of 1.5–4.2 dB, greater than the pure tone intensity discrimination thresholds for the comparable conditions of Experiment 1. Mean thresholds across all four conditions are 3.0 dB at 10-dB SL, 3.1 dB at 20-dB SL, and 2.1 dB at 30-dB SL, consistent with a small (~1-dB) improvement in performance with increasing standard level.

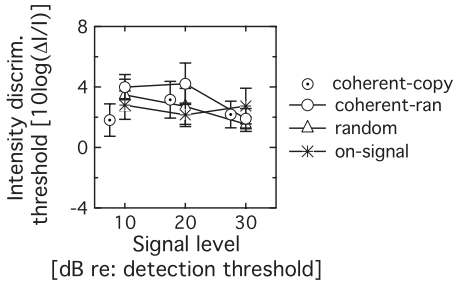


FIG. 2. Mean intensity discrimination thresholds for a narrowband noise signal are plotted as a function of standard level relative to detection threshold. Error bars indicate ±1 standard error of the mean across the six observers. Symbols reflect the masker condition, as indicated in the legend. The dotted circles, corresponding to the *coherent-copy* condition, where the signal is an exact copy of the on-signal masker to which it is added, have been shifted leftward on the abscissa to improve resolution of points.

These discrimination data were analyzed in two stages. The first stage assessed the pattern of results for the *coherent-ran*, *on-signal*, and *random* conditions. The *coherent-ran* condition was selected for this analysis for two reasons. First, the masking release was greater in this condition as compared to the *coherent-copy* condition, so any effects associated with masking release would plausibly be larger for this condition. Second, addition of the signal in this condition introduces across-frequency envelope differences, as it does for the pure tone stimulus used in Experiment 1; it was reasoned that an analysis of masker effects using this masking release condition would therefore be more comparable to previous results. A repeated-measures ANOVA was performed with three levels of MASKER (*coherent-ran*, *on-signal*, and *random*) and three levels of STANDARD (10-, 20-, and 30-dB SL). This analysis resulted in a main effect of STANDARD ($F_{2,10}=6.13, p<0.05$), no effect of MASKER ($F_{2,10}=2.59, p=0.12$), and no interaction ($F_{4,20}=1.25, p=0.32$). A preplanned contrast comparing the effect of standard level on intensity discrimination thresholds in the *coherent-ran* condition as compared to the *on-signal* and *random* conditions was not significant ($F_{1,5}=0.85, p=0.40$). These findings are consistent with the summary statement that thresholds tend to improve modestly with increasing stimulus level, and that this trend does not differ reliably across conditions associated with masking release (*coherent-ran*) and baseline conditions (*on-signal* and *random*).

The second analysis performed on the intensity discrimination data from Experiment 2 was a repeated-measures ANOVA with two levels of MASKER (*coherent-ran* and *coherent-copy*) and three levels of STANDARD (10-, 20-, and 30-dB SL). This analysis resulted in a main effect of MASKER ($F_{1,5}=12.67, p<0.05$). Both the main effect of STANDARD ($F_{2,10}=2.91, p=0.10$) and the interaction ($F_{2,10}=2.92, p=0.10$) failed to reach significance. This outcome indicates that whereas intensity discrimination thresholds were lower in the *coherent-copy* as compared to the *coherent-ran* condition, there was no statistical evidence of a differential effect of signal level across these conditions.

D. Discussion

Signal detection thresholds in the *on-signal* and *random* baseline conditions were elevated relative to those observed in comparable conditions of Experiment 1. This is consistent with the idea that observers were using features of the envelope statistics to detect a pure tone added to a narrowband of noise in the previous experiment. For example, Richards (2002) argued that both increased overall intensity and envelope flattening of the summed stimulus can contribute to sensitivity to a pure tone in a narrowband noise masker. Also in contrast to the results of Experiment 1, thresholds in these two baseline conditions differed significantly, with poorer performance in the *random* as compared to the *on-signal* conditions. While the reason for this difference is not clear, it suggests that there may be greater energetic masking or across-channel masking (Moore *et al.*, 1990; Buss, 2008) due to the flanking bands for a narrowband noise as compared to a tonal signal, perhaps due to the availability of envelope cues in tone detection.

Coherent masker envelope fluctuation improved detection relative to baseline thresholds for both the *coherent-copy* and *coherent-ran* conditions, with greater masking release for the random noise-band signal. This could be related to the fact that addition of the narrowband noise signal introduced an across-frequency envelope difference in the *coherent-ran* but not the *coherent-copy* conditions. In the latter condition the signal increased the relative level of the band at the signal frequency, but that band was still comodulated with the flanking maskers. Whereas the masking release observed under conditions of coherent masker fluctuation is often discussed in terms of the across-frequency envelope decorrelation associated with addition of a signal (Richards, 1987; van de Par and Kohlrausch, 1998), there is also precedent in the literature for obtaining a masking release in the absence of across-channel envelope decorrelation (Green and Nguyen, 1988; Hall and Grose, 1988).

Perhaps the most interesting aspect of the present data is the relative lack of an effect of standard level on intensity discrimination across all four masker conditions. Thresholds in Experiment 2 improved approximately 1 dB between 10 and 30 dB SL; in contrast, comparable pure tone data of Experiment 1 indicate an improvement of 2.4 dB in baseline conditions and 5.4 dB in masking release conditions. The finding of comparable intensity discrimination in baseline and masking release conditions for the narrowband noise signal indicates that masking release is not always associated with elevated suprathreshold discrimination thresholds. Whereas detection thresholds were 4.3 dB higher in the *random* than the *on-signal* masker conditions, consistent with masking associated with the introduction of flanking bands, intensity discrimination was not affected by the presence of flanking maskers. This finding fails to support the idea that masking more adversely affects intensity discrimination than detection.

The results of Experiment 2 are consistent with the hypothesis that the improvement observed with increasing level of a pure tone standard is due primarily to flattening of the temporal envelope of the summed stimulus, with reductions in envelope fluctuations supporting greater sensitivity to intensity changes across listening intervals. The finding of a small improvement in thresholds with increasing standard level of the narrowband noise signal could reflect a modest additional beneficial effect of spread of excitation. These results are also broadly consistent with the idea that stimulus fluctuation disrupts memory for intensity, an effect that may be more pronounced when the masker and signal-plus-masker are perceptually similar.

V. EXPERIMENT 3

The results of Experiment 2 are consistent with the idea that stimulus amplitude fluctuation plays an important role in the ability to discriminate intensity of a suprathreshold signal. In that paradigm the envelope modulation depth of the on-signal masker band summed with the signal itself does not depend on the SNR; for these stimuli, fluctuations conform to the envelope statistics of a 20-Hz band of Gaussian noise for all SNRs. The relatively poor intensity discrimination thresholds in all conditions of Experiment 2 (see Fig. 2) are consistent with the idea that performance is poor when the standard is characterized by marked envelope fluctuation, even as the level of the standard increases. Despite this, there was a slight but significant improvement in intensity discrimination as a function of level, an effect of about 1-dB improvement in threshold with a 20-dB increase in the standard level. This result suggests that absolute level could play a small but significant role in performance under conditions of pronounced stimulus fluctuation. The third experiment was designed to further assess the role of envelope fluctuation and absolute level in the improved intensity discrimination with increasing level of a pure tone standard, such as that observed in the data of Experiment 1.

Increasing the SNR of a pure tone signal in a narrowband noise has at least two effects: it tends to reduce inherent fluctuations of the signal/masker sum, and it also increases the opportunity to benefit from representation of the signal in multiple auditory channels due to spread of excitation. These two effects might not be mutually exclusive. Whereas intensity discrimination thresholds for stationary stimuli are likely to be limited by internal noise, thresholds for stimuli with fluctuating envelopes might be limited by external noise (Buss *et al.*, 2006). If the near miss to Weber's law is due in part to combination of information across auditory channels with independent internal noise (Florentine and Buus, 1981),

then the effect of stimulus level on performance would be expected to depend on the degree to which internal noise (as opposed to external noise) limits performance. In this context, reduced envelope fluctuation associated with increased SNR of a pure tone signal could improve performance by reducing external noise, with this reduction improving on-frequency cues and facilitating benefit derived from spread of excitation. If factors related to the reduction in envelope fluctuation and spread of excitation contribute synergistically to level effects in suprathreshold pure tone intensity discrimination, then spread of excitation would have a smaller effect in conditions where increasing level of the standard is *not* associated with reduced fluctuation. This would be consistent with the idea that the standard level effect observed in Experiment 2 was small because large amplitude fluctuations in the signal-plus-masker precluded taking full advantage of the detection benefits associated with spread of excitation.

The approach taken in Experiment 3 was to dissociate the two effects of increasing the SNR of a tonal standard added to narrowband noise. Whereas Experiment 2 incorporated highly fluctuating stimuli at a range of standard levels, Experiment 3 included stimuli with a range of envelope statistics, either with or without associated standard level increments. This approach allows a test of the hypothesis that the level effects for intensity discrimination of a tone in narrowband noise are the consequence of both absolute level effects and reduction in envelope fluctuation with increasing SNR. The procedures used to dissociate envelope and level effects of increasing SNR differ from those in previous experiments in several important respects. In conditions for which absolute level was held constant across SNR, the tonal standard and the narrowband noise masker at the signal frequency were summed and then that sum was scaled back to 50 dB SPL, the level of the *on-signal* masker alone. Another important procedural difference is that the standard and standard-plus-increment intervals differed only in the level of the summed stimulus: both the standard and the standard-plus-increment intervals contained a composite stimulus composed of a 1000-Hz tone and a narrowband masker centered on 1000 Hz, and the SNR of this composite stimulus was held constant across all intervals. The composite stimulus was gated on only during the listening intervals, analogous to the gating imposed on the pure tone standard alone in Experiment 1. In the *random* condition flanking bands were presented continuously.

Previous work has shown that asynchronous onset of maskers distributed across frequency can substantially disrupt processing characteristic of CMR (Dau *et al.*, 2004; Grose *et al.*, 2009). For that reason masker conditions in Experiment 3 were restricted to the *on-signal* and *random* masker conditions. The extent to which results in the baseline conditions generalize to coherent masker conditions will be addressed in the discussion, where data from Experiments 1 and 3 are compared.

A. Observers

Five observers participated in this experiment, including Obs 7 and Obs 12–15. There were two males in this group, and the mean age was 31 years.

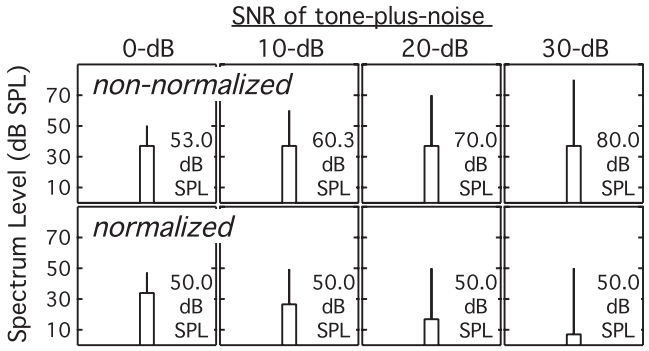


FIG. 3. Standard stimuli used in Experiment 3 are illustrated for both the *non-normalized* and *normalized* conditions and for the four values of SNR tested in each condition. The total level of the tone-plus-noise appears at the lower right of each panel.

B. Stimuli

Stimuli were based on those used in Experiment 1. The masker was either a single 20-Hz band of noise centered on 1000 Hz (*on-signal*) or a set of five bands of independent Gaussian noise centered on 200, 600, 1000, 1400, and 1800 Hz (*random*). The present experiment did not include a coherent masker fluctuation condition. The masker band centered on 1000 Hz was gated on during the listening intervals, with 50-ms raised-cosine ramps and a half-rise duration of 400 ms. In contrast, in the *random* conditions the flanking masker bands above and below the 1000-Hz frequency region played continuously. This gating manipulation was introduced to facilitate segregation of the tone and *on-signal* masker from the ongoing stream of flanking masker bands, thereby increasing confidence that any effect of flanking masker bands would be due to energetic masking as opposed to a failure to selectively attend to stimuli in the region of 1000 Hz; this manipulation has been shown to improve intensity discrimination under conditions for which best performance is supported by information in a restricted frequency region of the stimulus (e.g., Buss, 2008).

In all conditions both the standard and standard-plus-increment stimuli were generated as the sum of a 1000-Hz pure tone and a narrowband noise at the same center frequency, with SNRs of 0, 10, 20, or 30 dB. In one set of conditions the narrowband noise centered on 1000 Hz was 50 dB SPL, and the tone was 50, 60, 70, or 80 dB SPL. In a second set of conditions the composite stimuli with SNRs of 0, 10, 20, or 30 dB were scaled to a total level of 50-dB SPL in the standard intervals. These scaled stimulus conditions will be referred to as *normalized*. Idealized long-term power spectra of stimuli in the standard (no increment) intervals for these conditions appear in Fig. 3, with the total level of that portion of the stimulus centered on 1000 Hz in standard interval indicated in the lower right of each panel.

In both *normalized* and *non-normalized* conditions, the stimuli associated with standard and standard-plus-increment intervals were generated using identical procedures except that the composite stimulus was more intense in the standard-plus-increment interval. In neither case did the SNR differ across standard and standard-plus-increment intervals. These procedures allowed strict control of envelope fluctua-

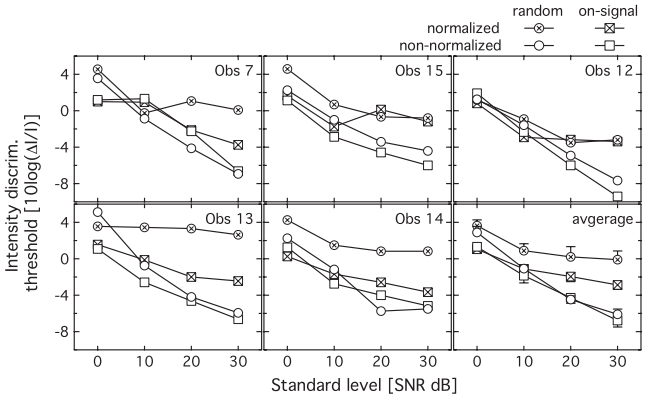


FIG. 4. Thresholds are plotted as a function of SNR, with a separate panel for each observer. The lower right panel shows the mean across observers, with error bars indicating ±1 standard error of the mean. Symbols reflect the masker conditions, as indicated in the legend.

tion statistics across conditions and across intervals within a condition, and ensured that factors related to the detection of a change in envelope statistics did not affect intensity discrimination. As in previous experiments, intensity discrimination thresholds are reported in units of $10\log(\Delta I/I)$. However, in the present experiment, intensity of the standard (I) and the intensity increment (ΔI) were computed based on the level of composite stimulus (signal plus 1000-Hz masker band) rather than on the pure tone alone.

C. Results

Performance varied across individuals, so thresholds for each observer are shown in Fig. 4. Thresholds are plotted in units of $10\log(\Delta I/I)$ for a range of SNRs, and symbols reflect stimulus condition as indicated in the legend. The standard error of the mean across replicate threshold estimates within observers had a median value of 0.56 dB. The standard error of the mean across the five observers, illustrated with error bars in the bottom right panel of Fig. 4, had a median value of 0.62 dB. Initial data for Obs 7 (not shown) indicated relatively stable thresholds across conditions and across standard levels, with all thresholds falling between −2.1 and 2.0 dB. This observer was given the opportunity to practice the task, and data collection was repeated; this second set of data, shown in Fig. 4, much more closely resembles the results of the other four observers.

For most observers and in most conditions, intensity discrimination thresholds improved with increasing SNR. In the *normalized* conditions this improvement can be attributed to a reduction in amplitude fluctuation in the output of auditory filters centered at the signal frequency. In the *non-normalized* conditions there is an additional potential effect of spread of excitation due to an increased opportunity to incorporate information from a wider range of off-frequency channels. Looking across observers, thresholds tended to fall in a roughly parallel fashion in all conditions between SNRs of 0 and 10 dB. Mean thresholds improved by an average of approximately 2.4 dB in the *normalized* conditions and 3.6 dB in the *non-normalized* conditions. This result is consistent with a marked benefit of a reduction in stimulus fluctuation and little additional benefit from off-frequency cues for this range of standard levels. For SNRs greater than 10 dB, there tends to be modest additional improvement in thresholds with increased SNR in *normalized* conditions, with average thresholds improving 1.4 dB between 10 and 30n dB SNR. In contrast, thresholds in the *non-normalized* condition continued to improve another 5.0 dB on average with further increases in SNR.

These observations of the data were assessed statistically with a repeated-measures ANOVA, with two levels of MASKER (*random* and *on-signal*), two levels of CONDITION (*normalized* and *non-normalized*), and four levels of SNR (0–30 in 10-dB steps). Significant main effects included MASKER ($F_{1,4}=15.08, p<0.05$), CONDITION ($F_{1,4}=63.04, p<0.01$), and SNR ($F_{3,12}=130.46, p<0.0001$). The CONDITION-by-SNR interaction was also significant ($F_{3,12}=23.84, p<0.0001$), but no other interaction approached significance ($p>0.10$). This result supports the observation that signal level has differential effects in the normalized and non-normalized conditions.

These results are broadly consistent with the conclusion that envelope fluctuation limits performance at low SNRs, and benefits related to spread of excitation play a role primarily at SNRs above 10 dB, where stimulus fluctuation (i.e., external noise) imposes less of a limit to performance. However, there appear to be notable individual differences in the ability to use these cues. One aspect of individual differences in these data is seen in the relationship between *on-signal* and *random* thresholds in the *normalized* stimulus conditions (filled symbols in Fig. 4). For some observers thresholds are similar in the *normalized/random* and *normalized/on-signal* conditions (e.g., Obs 12 and 15), whereas for others thresholds are consistently 2–6 dB poorer in the *normalized/random* than the *normalized/on-signal* condition (e.g., Obs 13 and 14). This difference across data sets could reflect greater susceptibility to off-frequency masking in some observers.

D. Discussion

The results of Experiment 3 are consistent with the conclusion that improved intensity discrimination with increasing level of the standard tone in Experiment 1 is dominated by reductions in amplitude fluctuation for low levels of the standard and with introduction of off-frequency cues related to spread of excitation at higher levels of the standard tone. There is sparse evidence of masking associated with the presence of flanking maskers. Overall, the thresholds were elevated 2.4 dB by the presence of random sidebands in *normalized* conditions and 0.7 dB in *non-normalized* conditions. The fact that this effect is level dependent, with slightly smaller effects in the *non-normalized* condition, is consistent with published data for off-frequency masking in intensity discrimination. Greenwood (1993) speculated that level effects for off-frequency masking could be due to the increased excitation associated with the standard "overcoming" excitation related to a neighboring masker, such that broad changes in excitation due to addition of the signal would not be fully masked. Interpretation of threshold elevation in the presence of random flanking bands in terms of energetic masking is

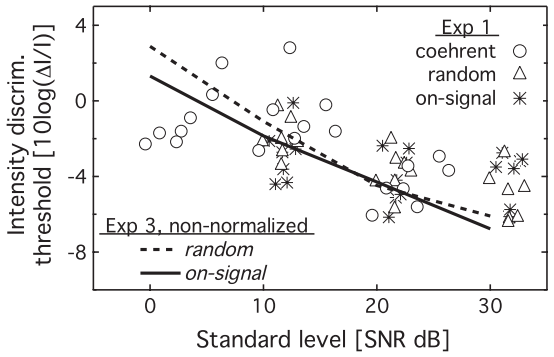


FIG. 5. Intensity discrimination thresholds in the narrowband noise conditions of Experiment 1 were recomputed relative to the level of the pure tone signal and the masker band to which it was added. The results are plotted for individual observers as a function of SNR, and symbols reflect the masker condition. The lines show mean results of the *non-normalized* conditions of Experiment 3.

undermined, however, by the finding of substantial individual differences as well as the failure to find a significant interaction between masker type (*on-signal* and *random*) and stimulus condition (*normalized* and *non-normalized*).

In contrast to Experiment 1, in the present paradigm flanking maskers were played continuously and the on-signal masker was gated on only during the listening interval, synchronously with the standard tone. This procedural difference was meant to highlight the tone and the masker in the 1000-Hz frequency region and to reduce possible confusion regarding which stimulus components are most relevant in making intensity discriminations across intervals. Another procedural difference was that the SNR was held constant across standard and standard-plus-increment intervals in Experiment 3, with intensity increments produced by a scalar applied to the tone-plus-masker composite. In order to evaluate the consistency of results obtained in these two experiments, thresholds were compared across paradigms in the following manner.

Figure 5 shows intensity discrimination as a function of standard level in dB SNR for individual observers in the three narrowband noise conditions of Experiment 1, as indicated by symbol shape. These thresholds were adjusted to incorporate the on-signal masker into the estimate of the standard level (I), similar to the approach taken in Experiment 3. This adjustment had a larger effect at the low than high SNRs, with mean reductions in threshold estimates of 2.0 dB for standard levels 0–5 dB, 0.3 dB for standard levels of 10–15 dB, and smaller effects at higher standard levels. Lines in Fig. 5 indicate mean thresholds in the *non-normalized/random* (dashed) and *non-normalized/on-signal* (solid) conditions of Experiment 3. The level effects observed in the *non-normalized* conditions of Experiment 3 capture the general trends in the data of Experiment 1, including thresholds in the *coherent* condition. Interpretation of this relationship is tempered by substantial individual differences, coupled with the fact that only one observer (Obs 7) provided data in both experiments. However, the general agreement across data sets is consistent with the conclusion that suprathreshold intensity discrimination performance in coherent masker conditions is relatively poor near detection threshold due to the detrimental effects of stimulus envelope fluctuation. There is little evidence that additional effects related to loss of signal information following across-channel comparisons (coarseness) or to the masking associated with flanking maskers.

Recall that one hypothesis regarding elevated thresholds in masking release conditions has to do with the masker fluctuation between listening intervals corrupting trace memory for the level of the tone in each listening interval. By this account, masker fluctuation between listening intervals is more disruptive for a tonal signal played at a low SNR, perhaps due to perceptual similarity between the masker and signal-plus-masker stimuli. Data for comparable continuous and gated masker conditions were collected for the on-signal masker in Experiments 1 and 3, respectively. While there are large individual differences, mean thresholds for the continuous and gated masker are quite close for the lower two signal levels. At the highest signal level thresholds for the continuous masker conditions (Fig. 5, stars) are approximately 2 dB greater than those in the analogous gated masker condition (Fig. 5, solid lines). It is unclear whether to attribute this difference to individual variability or the effect of fixing SNR across intervals or to a reliable difference between gated and continuous masker presentation, but in any either case this pattern of results fails to support an effect of masker variability in the intertrial interval via corruption of a trace memory for level, where the largest effects would be predicted for the low rather than high signal levels.

VI. GENERAL DISCUSSION AND CONCLUSIONS

The present set of experiments was carried out to further understand intensity discrimination under conditions of masking release due to masker envelope coherence. As previously demonstrated, suprathreshold intensity discrimination for pure tone signals presented near threshold was poorer under coherent masker conditions than in baseline masking conditions at comparable levels relative to detection threshold (dB SL). Experiment 1 showed that similar suprathreshold effects can be demonstrated for both narrowband and bandpass noise masking release paradigms: in the first case masker envelope coherence is based on inherent modulation of narrowband noise maskers, and in the second case it is based on multiplication with the envelope of an independent narrowband noise. Additional control conditions confirmed that this suprathreshold deficit was not dependent on absolute signal level in either masking release paradigm. Experiment 2 showed that intensity discrimination was uniformly poor irrespective of masker condition when the signal was a narrow band of noise rather than a pure tone, consistent with the interpretation that inherent fluctuation may limit intensity discrimination for an increment added to a fluctuating standard. The final experiment assessed the relative contribution of envelope modulation reduction and increasing spread of excitation in the finding of improved intensity discrimination thresholds with increasing SNR for a pure tone standard. It was also hypothesized that spread of excitation could improve performance, particularly in combination with reduced envelope fluctuation at higher SNRs. The

results of Experiment 3 were characterized by individual differences, but were broadly consistent with a beneficial effect of reduced envelope fluctuation for relatively low SNRs. Spread of excitation appeared to contribute primarily at the higher SNRs. The presence of random flanking maskers elevated thresholds by a few decibels in some observers. However, the nonuniformity of this effect across observers and the small average effect size suggest that energetic masking of upward spread of excitation plays a minor role in the pattern of suprathreshold intensity discrimination with coherent maskers.

Taken together, results are consistent with the conclusion that level fluctuation of the stimulus components at the signal frequency, the sum of the narrowband masker and the signal, interferes with intensity discrimination in a comparable fashion across masker conditions. This effect is most evident in the coherent modulation as compared to baseline conditions when results are plotted as a function of the SL of the standard. Plotting the results in absolute signal level or SNR, as in Fig. 5, illustrates the approximate uniformity of level effects across masker conditions.

The present data suggest that the disruptive effects of inherent stimulus fluctuation within the listening interval is the most parsimonious explanation for the poor suprathreshold intensity discrimination performance observed in the presence of comodulated narrowband noise maskers, both in the present experiments and in the published data (Hall and Grose, 1995). There was no indication that derived cues based on across-frequency comparisons were less informative regarding intensity of the signal than cues in the baseline conditions. Whereas flanking maskers may elevate detection thresholds, particularly in Experiment 2, there was little evidence that masking is responsible for the reduced masking release for discrimination. Minimal data on gated as compared to continuous presentation of an on-signal masker presented alone cast doubt on the idea that masker fluctuation between listening intervals plays a role in the present results. Results of the final experiment indicate that envelope stimulus fluctuation associated with increasing SNR of a pure tone signal may reduce thresholds by improving the quality of cues at the signal frequency and by increasing the ability to benefit from spread of excitation, both effects related to reduced external noise.

It is interesting to speculate that similar factors could be responsible for the poor suprathreshold intensity discrimination observed under conditions of monaural and binaural masking release (Henning, 1991). Stimuli composed of a tonal standard in noise would be associated with greater envelope fluctuation at low than high SNR for binaural as well as monaural presentation. It is also possible that increased external noise associated with stimulus fluctuation could contribute to the finding of relatively poor gap detection for a tonal carrier presented at a low SNR in a narrowband noise background (Hall and Grose, 1992). As in intensity discrimination, stimulus envelope fluctuation is associated with poor gap detection (Shailer and Moore, 1983; Eddins *et al.*, 1992). Results of the present experiments could also be related to the finding of relatively poor suprathreshold pitch ranking for tones presented in comodulated noise (Hall *et al.*, 1997). Perceived pitch is affected by stimulus level (for a review, see Jesteadt and Neff, 1982), so it is possible that envelope fluctuation of a tone-plus-masker could introduce variability in perceived pitch. This possibility is the topic of current research.

Reduced sensitivity in suprathreshold discrimination tasks for a signal masked by a coherently fluctuating noise is of theoretical interest in understanding basic psychoacoustic findings (e.g., CMR), but it may also be relevant the ability to process auditory stimuli under more natural listening conditions. In normal-hearing listeners, masking of a speech signal in noise can be reduced by the introduction of masker level fluctuation, with the biggest effects for relatively slow rates of modulation (Miller and Licklider, 1950; Bacon *et al.*, 1998). It has been argued that this result can be explained in terms of the reduced masker level in the modulation minima, associated with brief "glimpses" of the signal at an improved SNR (Dirks and Bower, 1970). Masker fluctuation is not as beneficial for listeners with moderate sensorineural hearing impairment as it is for normal hearing listeners (Festen and Plomp, 1990), even when controlling for the effects of audibility (Eisenberg *et al.*, 1995). Poorer temporal resolution and/or frequency selectivity in hearing-impaired listeners have been suggested to account for this result (Festen and Plomp, 1990; Baer and Moore, 1994; Eisenberg *et al.*, 1995; Bacon *et al.*, 1998), but the factors responsible for poor ability to benefit from masker level fluctuations in cochlear hearing loss are still unknown. Results of the present study with normal-hearing listeners indicate that suprathreshold intensity discrimination could also play a role in this finding.

The poor suprathreshold speech perception in amplitude modulated noise demonstrated by Grose and Hall (1992) could be affected by the fidelity with which intensity cues for speech are encoded in modulated noise. It is also likely that suprathreshold pitch discrimination and temporal processing of speech cues could limit performance on speech recognition tasks in fluctuating noise. Whereas the finding of masking release for both coherent and incoherent modulations across frequency indicates that the masking release for speech may not be closely allied with CMR (Howard-Jones and Rosen, 1993), the findings related to stimulus fluctuation at low SNRs could also apply to a wide range of conditions associated with masking release, not just those described in the CMR literature. More work is needed to assess the possible role of stimulus fluctuation and suprathreshold intensity discrimination in the perception of speech in modulated noise.

ACKNOWLEDGMENTS

This work was supported by a grant from NIH NIDCD (Grant No. R01 DC007391). John Grose and two anonymous reviewers provided helpful comments on this manuscript.

[1]Here the phrase "suprathreshold" refers to the level of the standard. This use is in contrast to the work of Wojtczak and Viemeister (2008), who studied perception of suprathreshold changes in intensity, where the changes themselves were suprathreshold.

ANSI (**1996**). ANSI S3-1996, American National Standards Specification

for Audiometers (American National Standards Institute, New York).
Bacon, S. P., Opie, J. M., and Montoya, D. Y. (**1998**). "The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds," J. Speech Lang. Hear. Res. **41**, 549–563.
Baer, T., and Moore, B. C. J. (**1994**). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," J. Acoust. Soc. Am. **95**, 2277–2280.
Berg, B. G. (**1996**). "On the relation between comodulation masking release and temporal modulation transfer functions," J. Acoust. Soc. Am. **100**, 1013–1023.
Bos, C. E., and de Boer, E. (**1966**). "Masking and discrimination," J. Acoust. Soc. Am. **39**, 708–715.
Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., and Purks, S. R. (**1984**). "Intensity perception. XIII. Perceptual anchor model of context-coding," J. Acoust. Soc. Am. **76**, 722–731.
Buss, E. (**2008**). "The effect of masker level uncertainty on intensity discrimination," J. Acoust. Soc. Am. **123**, 254–264.
Buss, E., and Hall, J. W. (**2008**). "Factors contributing to comodulation masking release with dichotic maskers," J. Acoust. Soc. Am. **124**, 1905–1908.
Buss, E., Hall, J. W., and Grose, J. H. (**2006**). "Development and the role of internal noise in detection and discrimination thresholds with narrow band stimuli," J. Acoust. Soc. Am. **120**, 2777–2788.
Buus, S. (**1985**). "Release from masking caused by envelope fluctuations," J. Acoust. Soc. Am. **78**, 1958–1965.
Carlyon, R. P., and Beveridge, H. A. (**1993**). "Effects of forward masking on intensity discrimination, frequency discrimination, and the detection of tones in noise," J. Acoust. Soc. Am. **93**, 2886–2895.
Dau, T., Ewert, S. D., and Oxenham, A. J. (**2004**). in *Auditory Signal Processing: Physiology, Psychoacoustics and Models*, edited by D. Pressnitzer, A. de Cheveigne, S. McAdams, and L. Collet (Springer, New York), pp. 335–343.
Dirks, D. D., and Bower, D. (**1970**). "Effect of forward and backward masking on speech intelligibility," J. Acoust. Soc. Am. **47**, 1003–1008.
Eddins, D. A. (**2001**). "Monaural masking release in random-phase and low-noise noise," J. Acoust. Soc. Am. **109**, 1538–1549.
Eddins, D. A., Hall, J. W. III, and Grose, J. H. (**1992**). "The detection of temporal gaps as a function of frequency region and absolute noise bandwidth," J. Acoust. Soc. Am. **91**, 1069–1077.
Eisenberg, L. S., Dirks, D. D., and Bell, T. S. (**1995**). "Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing," J. Speech Hear. Res. **38**, 222–233.
Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.
Florentine, M., and Buus, S. (**1981**). "An excitation-pattern model for intensity discrimination," J. Acoust. Soc. Am. **70**, 1646–1654.
Green, D. M., and Nguyen, Q. T. (**1988**). "Profile analysis: Detecting dynamic spectral changes," Hear. Res. **32**, 147–163.
Greenwood, D. D. (**1993**). "The intensive DL of tones: Dependence of signal/masker ratio on tone level and on spectrum of added noise," Hear. Res. **65**, 1–39.
Grose, J. H., Buss, E., and Hall, J. W. (**2009**). "Within- and across-channel factors in the multiband comodulation masking release paradigm," J. Acoust. Soc. Am. **125**, 282–293.
Grose, J. H., and Hall, J. W. (**1992**). "Comodulation masking release for speech stimuli," J. Acoust. Soc. Am. **91**, 1042–1050.
Haggard, M. P., Hall, J. W., and Grose, J. H. (**1990**). "Comodulation masking release as a function of bandwidth and test frequency," J. Acoust. Soc. Am. **88**, 113–118.
Hall, J. W., and Grose, J. H. (**1988**). "Comodulation masking release: Evidence for multiple cues," J. Acoust. Soc. Am. **84**, 1669–1675.
Hall, J. W., and Grose, J. H. (**1992**). "Masking release for gap detection," Philos. Trans. R. Soc. London, Ser. B **336**, 331–337.
Hall, J. W., and Grose, J. H. (**1995**). "Amplitude discrimination in masking release paradigms," J. Acoust. Soc. Am. **98**, 847–852.
Hall, J. W., Grose, J. H., and Dev, M. B. (**1997**). "Signal detection and pitch ranking in conditions of masking release," J. Acoust. Soc. Am. **102**, 1746–1754.
Hall, J. W., Grose, J. H., and Moore, B. C. J. (**1993**). "Influence of frequency selectivity on comodulation masking release in normal-hearing listeners," J. Speech Hear. Res. **36**, 410–423.
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.
Henning, G. B. (**1991**). in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 507–512.
Howard-Jones, P. A., and Rosen, S. (**1993**). "Uncomodulated glimpsing in 'checkerboard' noise," J. Acoust. Soc. Am. **93**, 2915–2922.
Jesteadt, W., and Neff, D. L. (**1982**). "A signal-detection-theory measure of pitch shifts in sinusoids as a function of intensity," J. Acoust. Soc. Am. **72**, 1812–1820.
Jesteadt, W., Nizami, L., and Schairer, K. S. (**2003**). "A measure of internal noise based on sample discrimination," J. Acoust. Soc. Am. **114**, 2147–2157.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Miller, G. A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173.
Moore, B. C. J., Glasberg, B. R., and Schooneveldt, G. P. (**1990**). "Across-channel masking and comodulation masking release," J. Acoust. Soc. Am. **87**, 1683–1694.
Moore, B. C. J., Shailer, M. J., Hall, J. W., and Schooneveldt, G. P. (**1993**). "Comodulation masking release in subjects with unilateral and bilateral hearing impairment," J. Acoust. Soc. Am. **93**, 435–451.
Oberfeld, D. (**2008**). "The mid-difference hump in forward-masked intensity discrimination," J. Acoust. Soc. Am. **123**, 1571–1581.
Plack, C. J. (**1996**). "Loudness enhancement and intensity discrimination under forward and backward masking," J. Acoust. Soc. Am. **100**, 1024–1030.
Plack, C. J., Carlyon, R. P., and Viemeister, N. F. (**1995**). "Intensity discrimination under forward and backward masking: Role of referential coding," J. Acoust. Soc. Am. **97**, 1141–1149.
Richards, V. M. (**1987**). "Monaural envelope correlation perception," J. Acoust. Soc. Am. **82**, 1621–1630.
Richards, V. M. (**2002**). "Varying feedback to evaluate detection strategies: The detection of a tone added to noise," J. Assoc. Res. Otolaryngol. **3**, 209–221.
Schooneveldt, G. P., and Moore, B. C. (**1987**). "Comodulation masking release (CMR): Effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band," J. Acoust. Soc. Am. **82**, 1944–1956.
Shailer, M. J., and Moore, B. C. (**1983**). "Gap detection as a function of frequency, bandwidth, and level," J. Acoust. Soc. Am. **74**, 467–473.
Spiegel, M. F., Picardi, M. C., and Green, D. M. (**1981**). "Signal and masker uncertainty in intensity discrimination," J. Acoust. Soc. Am. **70**, 1015–1019.
Swets, J. A., Shipley, E. F., McKey, M. J., and Green, D. M. (**1959**). "Multiple observations of signals in noise," J. Acoust. Soc. Am. **31**, 514–521.
van de Par, S., and Kohlrausch, A. (**1998**). "Comparison of monaural (CMR) and binaural (BMLD) masking release," J. Acoust. Soc. Am. **103**, 1573–1579.
Wojtczak, M., and Viemeister, N. F. (**2008**). "Perception of suprathreshold amplitude modulation and intensity increments: Weber's law revisited," J. Acoust. Soc. Am. **123**, 2220–2236.
Zeng, F. G. (**1998**). "Interactions of forward and simultaneous masking in intensity discrimination," J. Acoust. Soc. Am. **103**, 2021–2030.
Zeng, F. G., Turner, C. W., and Relkin, E. M. (**1991**). "Recovery from prior stimulation. II: Effects upon intensity discrimination," Hear. Res. **55**, 223–230.
Zwicker, E. (**1970**). "Masking and psychological excitation as consequences of the ear's frequency analysis," in *Frequency Analysis and Periodicity Detection in Hearing. The Proceedings of the International Symposium on Frequency Analysis and Periodicity Detection in Hearing*, Driebergen, The Netherlands, 23–27 June **1969**, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden).

Combination of masking releases for different center frequencies and masker amplitude statistics

Bastian Epp[a] and Jesko L. Verhey
Graduate School Neurosensory Science and Systems, Institut für Physik, Carl von Ossietzky Universität Oldenburg, Carl-von-Ossietzky-Straβe 9-11, 26111 Oldenburg, Germany

(Received 2 December 2008; revised 14 July 2009; accepted 26 July 2009)

Several masking experiments have shown that the auditory system is able to use coherent envelope fluctuations of the masker across frequency within one ear as well as differences in interaural disparity between signal and masker to enhance signal detection. The two effects associated with these abilities are comodulation masking release (CMR) and binaural masking level difference (BMLD). The aim of the present study was to investigate the combination of CMR and BMLD. Thresholds for detecting a sinusoidal signal were measured in a flanking-band paradigm at three different signal frequencies. The masker was presented diotically, and various interaural phase differences (IPDs) of the signal were used. The masker components were either multiplied or Gaussian narrowband noises. In addition, a transposed stimulus was used to increase the BMLD at a high signal frequency. For all frequencies and masker conditions, thresholds decreased as the signal IPD increased and were lower when the masker components were comodulated. The data show an addition of the monaural and binaural masking releases in decibels when masker conditions with and without comodulation and the same spectrum were compared.
 [DOI: 10.1121/1.3205404]

PACS number(s): 43.66.Dc, 43.66.Rq, 43.66.Mk, 43.66.Ba [BCM] Pages: 2479–2489

I. INTRODUCTION

An important task of the auditory system in natural acoustical environments is to segregate sounds from different sound sources. It is generally assumed that the auditory system uses monaural cues, such as the coherent envelope fluctuations across frequency—a characteristic of many natural sounds (Nelken *et al.*, 1999)—as well as binaural cues to separate the different sound sources. The aim of the present study was to investigate the ability of the auditory system to combine binaural and monaural cues in a psychoacoustical masking paradigm.

One psychoacoustical phenomenon associated with the ability to use monaural across-frequency cues is comodulation masking release (CMR). CMR is the effect that the detectability of a sinusoid masked by a narrowband noise centered at the signal frequency [signal-centered band (SCB)] can be improved by additional masker bands at spectrally distal positions [commonly referred to as flanking bands (FBs)], but only if the SCB and the FBs show coherent envelope fluctuations, i.e., are comodulated (Hall *et al.*, 1984; Verhey *et al.*, 2003, for a review). For the FB paradigm, the magnitude of CMR is either calculated as the difference in thresholds with the SCB only (reference condition, RF) and the threshold obtained by addition of comodulated FBs (comodulated condition, CM), or is defined as the benefit due to comodulated noise bands compared to noise bands having uncorrelated intensity fluctuations (uncorrelated condition, UN). In the following, the former CMR will be referred to as CMR(RF-CM) and the latter will be referred to as CMR(UN-CM). CMR has been shown to depend on center frequency, number, spectral width and level of the FBs, and the statistics of the masker (Schooneveldt and Moore, 1987; Hall *et al.*, 1990; Moore *et al.*, 1990; Verhey *et al.*, 2007; Epp and Verhey, 2009). CMR tends to increase with signal frequency and with the number of FBs. In addition, CMR depends on the masker envelope distribution (Eddins, 2001; Epp and Verhey, 2009). For spectral configurations with FBs close to the signal frequency, it has been suggested that part of the CMR is due to processing the output of one auditory filter (e.g., McFadden, 1986; Schooneveldt and Moore, 1987; Piechowiak *et al.*, 2007). It was shown by Verhey *et al.* (1999) that a model of the auditory system can predict CMR in one type of CMR experiment, the bandwidening experiment, by exclusively processing the information at the output of the auditory filter centered at the signal frequency. In bandwidening CMR experiments, FBs are added implicitly by broadening the masker centered at the signal frequency. Piechowiak *et al.* (2007) showed that the model proposed by Verhey *et al.* (1999) also predicts the CMR in FB CMR experiments, i.e., where a FB was added at a spectrally distal position, if a moderate spectral distance to the SCB was used (less than 30% of the signal frequency). For a FB more distal to the signal frequency, CMR is assumed to be the result of across-channel processing (Cohen, 1991; Verhey *et al.*, 2003), at least if the SCB and the FB have the same level. However, Ernst and Verhey (2006) showed that, for large level differences between the SCB and the FB, part of the CMR might be due to suppression at a cochlear level, even when the FB center frequency is several octaves below the signal frequency.

The auditory system is also able to use interaural disparities in either the signal or the masker to improve the

[a]Author to whom correspondence should be addressed. Electronic mail: bastian.epp@uni-oldenburg.de

detectability of the masked signal (Hirsh, 1948; Licklider, 1948). This effect is referred to as the binaural masking level difference (BMLD) (Jeffress *et al.*, 1956; van de Par and Kohlrausch, 1999). The BMLD depends, among other things, on the bandwidth of the masker and the signal frequency (Hirsh, 1948; Zurek and Durlach, 1987; van de Par and Kohlrausch, 1999). The BMLD decreases as the masker bandwidth increases and has a tendency to decrease with the signal frequency. The question whether the auditory system is able to combine the two cues comodulation and interaural disparities to increase the efficiency of signal detection in noise was addressed by Hall *et al.* (1988, 2006) and Cohen and Schubert (1991). Hall *et al.* (1988) investigated CMR when the 500-Hz pure tone signal was masked by a narrow band of noise alone and when a comodulated FB was added at a center frequency of 400 Hz. The CMR was found to be larger in conditions with diotic stimulation than in conditions where an antiphasic signal was masked by a diotic noise. All of the subjects showed a reduced or no benefit due to the addition of a comodulated masker band. About half of the subjects could benefit from the comodulated FB, but only for the smallest bandwidth of the noise. For the largest bandwidth used in their study, Hall *et al.* (1988) concluded that the auditory system does not seem to be able to benefit from across-frequency information in a dichotic listening condition. Cohen and Schubert (1991) measured thresholds for a 700-Hz pure tone signal masked by a narrowband masker centered at the signal frequency in the presence and absence of an additional narrowband masker centered at 600 Hz. The stimuli were presented diotically (N_0S_0) and dichotically with an interaurally inverted signal (N_0S_π). They found a reduction in both CMR(RF-CM) and CMR(UN-CM) for the dichotic listening condition compared to the diotic condition. CMR(RF-CM) vanished but a small benefit was observed in the presence of a comodulated FB compared to an uncorrelated FB. As a small CMR(UN-CM) was found, Cohen and Schubert (1991) concluded that CMR and BMLD "are, to some extent, additive." In a more recent study of Hall *et al.* (2006), the combination of CMR and BMLD was investigated by changing the interaural correlation of the masker. The stimulus used in their study was a 500-Hz pure tone masked by several narrow bands of noise with a spectral distance of 100 Hz. They found no consistent binaural CMR, i.e., large individual differences and, on average, only a small enhancement of binaural detection due to the presence of the comodulated FB.

Schooneveldt and Moore (1989) also investigated the combination of CMR and BMLD using various frequency separations of SCB and FB and various monaural and binaural presentations of the masker and signal. The binaural benefit was quantified by comparison of conditions with diotic noise and monaural signal presentation (N_0S_m) and conditions with diotic noise and diotic signal presentation (N_0S_0). They hypothesized sequential processes underlying CMR and BMLD. This study is not directly comparable to the other studies because they investigated the combination of CMR and BMLD only for a binaural gain due to a comparison of a monaural versus diotic signal presentation. In a recent study, Epp and Verhey (2009) showed that data from a combined CMR(UN-CM) and BMLD paradigm at 700 Hz using various interaural phase differences (IPDs) of the signal in combination with diotically presented uncorrelated and comodulated masker conditions can be explained using a model with serial alignment of across-frequency and across-ear processing stages. This result supports the hypothesis of Schooneveldt and Moore (1989) that the processing stages underlying CMR and BMLD operate sequentially. However, Epp and Verhey (2009) only investigated the combination of CMR(UN-CM) with the BMLD.

The previous studies on the combination of comodulation and IPD suffer from at least one of the following limitations: (i) they used small spectral distances between the masker components, so within-channel mechanisms may have contributed to the CMR to a large extent; (ii) they quantified CMR using only one of the two definitions of CMR; and (iii) there is only a very limited set of data for comparison of the two single effects CMR and BMLD and their combination.

The present study attempts to overcome the limitations of these previous studies by measuring the thresholds for detecting a signal in the reference, uncorrelated and comodulated conditions for (i) larger spectral distances between the components, (ii) various IPD of the signal, and (iii) different signal frequencies.

The combined masking release was measured for two different noise types which have commonly been used in CMR experiments. The use of the two noise types facilitates the comparison to previous data in the literature. These two noise types differ in their envelope amplitude distributions and thus may provide insights into the mechanisms underlying the across-frequency and across-ear processing.

II. GENERAL METHODS

A. Procedure

A three-alternative, forced-choice procedure with adaptive signal-level adjustment was used to determine the masked threshold of the sinusoidal signal. The intervals in a trial were separated by gaps of 500 ms. Subjects had to indicate which of the intervals contained the signal. Visual feedback was provided after each response. The signal level was adjusted according to a two-down, one-up rule to estimate the 70.7% point on the psychometric function (Levitt, 1971). The initial step size was 8 dB. After every second reversal, the step size was halved, until a step size of 1 dB was reached. The run was then continued for another six reversals. The mean level at these last six reversals was used as an estimate of the threshold. The final individual threshold estimate was taken as the mean over four threshold estimates.

B. Stimuli and apparatus

The signal was a pure tone which was temporally centered in the masker and had a duration of 250 ms, including 50-ms raised-cosine ramps at onset and offset. The masker duration was 500 ms, including 50-ms raised-cosine ramps at onset and offset. The masker consisted of one or five noise bands. Each noise band had a bandwidth of 24 Hz and a

level of 60-dB SPL. One noise band was centered at the signal frequency (SCB), and the four flanking noise bands were centered at frequencies remote from the signal frequency (FBs). The FBs were either absent (reference condition, RF), had uncorrelated intensity fluctuations (uncorrelated condition, UN), or had the same intensity fluctuations (comodulated condition, CM) as the SCB. The masker was presented diotically. The signal had an IPD in the range from 0° to 180°. Two types of masking noise were used: multiplied noise and Gaussian noise.

Multiplied noise masker bands were generated by multiplying a random phase sinusoidal carrier at the desired center frequency by a narrowband noise, which was lowpass filtered at 12 Hz and where the dc component was removed. This procedure mimics the analog realization of multiplied noise with noise generators that produced signals with zero mean value. A similar procedure was used in previous studies (Ernst and Verhey, 2006). For the reference (RF) and the uncorrelated (UN) conditions, independent realizations of the lowpass noise were used for each masker band. In the comodulated (CM) condition, the same lowpass noise was used for all masker components.

Gaussian-noise bands were generated in the frequency domain by assigning numbers derived from draws of a normally distributed process to the real and complex parts of the desired frequency components in each band. For the RF and the UN conditions this was done independently for each noise band, while the numbers of a single draw were assigned to all frequency bands for the CM condition. The real part of the subsequent inverse fast Fourier transform yielded the desired waveform.

For both noise types, new random numbers were drawn for each interval and each trial. All signals were generated digitally with a sampling frequency of 44 100 Hz using MATLAB. Signals were converted to the analog domain (RME ADI-8 DS), amplified (Tucker Davis Technologies HB7), and presented to the listeners in a double-walled sound-attenuating booth via headphones. The type of headphones differed between the experiments and is specified in the corresponding methods section.

C. Listeners

Nine listeners participated in each experiment, varying in age from 22 to 28 years (one of them being the first author, BE). None of the listeners had any history of hearing difficulties and their audiometric thresholds were 15-dB HL or less in the relevant frequency range from 125 to 8000 Hz. The listeners had at least 2-h experience in experiments on CMR and binaural experiments before collecting the data. The listeners were the same in the second and third experiments. One of the listeners who participated in the second and third experiments also participated in the first experiment (TK).

III. EXPERIMENT 1: CMR AND BMLD AT 700 Hz

A. Rationale

To facilitate the interpretation of the combined effect of comodulation and interaural disparities, a signal frequency was chosen at which monaural and binaural cues alone were expected to lead to a masking release of a reasonable magnitude (Hall *et al.*, 1990; van de Par and Kohlrausch, 1999). Interaural disparities were gradually introduced and systematically combined using the SCB alone, additional uncorrelated maskers, and comodulated maskers. The resulting overall release from masking for each cue combination was used to interpret the combined effect.

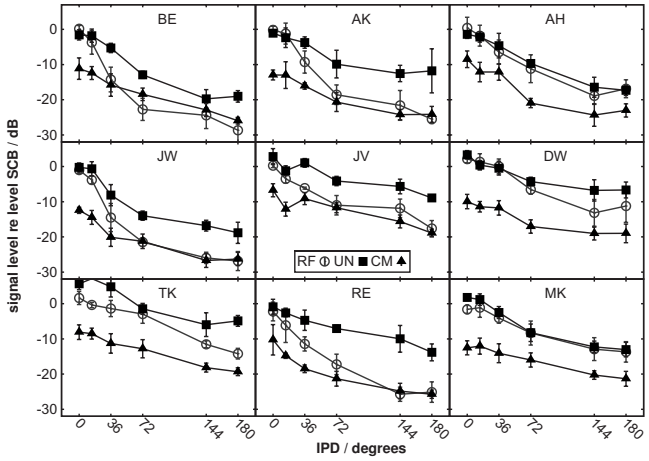


FIG. 1. Individual data for a signal frequency of 700 Hz and the multiplied noise masker. Mean thresholds are shown over four runs for the reference condition (RF, circles), the uncorrelated condition (UN, squares), and the comodulated condition (CM, triangles). Thresholds are plotted relative to the level of the masker centered at the signal frequency (SCB) as a function of the IPD. Error bars indicate ±1 standard deviation.

B. Methods

The signal and the SCB were located at 700 Hz. The FBs were located at 300, 400, 1000, and 1100 Hz. The signal IPD was 0° (diotic), 14.4°, 36°, 72°, 144°, or 180° (antiphasic). The stimuli were presented using Sennheiser HDA 200 audiometric headphones.

C. Results

Figure 1 shows individual results for the multiplied noise masker. Thresholds for detecting the signal in diotic conditions were highest for the reference (circles) and the uncorrelated (squares) masker conditions and lowest for the comodulated (triangles) condition. The thresholds for all listeners in all conditions decreased with increasing IPD. This means that for all masker conditions, an increase in the BMLD occurred as the IPD increased. The magnitude of the maximum BMLD (difference in threshold for the diotic and the antiphasic conditions) differed across the listeners. The maximum BMLD in the multi-band conditions (uncorrelated and comodulated conditions) varied from about 10 dB for listener DW to about 20 dB for listener BE. In the single-band condition (reference condition) the maximum BMLD varied from about 10 dB for listener MK to 28 dB for listener AK. There were also individual differences in the effect of the number of bands on the BMLD. For six of the listeners the BMLD was larger for the single-band condition than for the multi-band conditions. Such large individual differences have been reported before (Buss *et al.*, 2007). Three of the

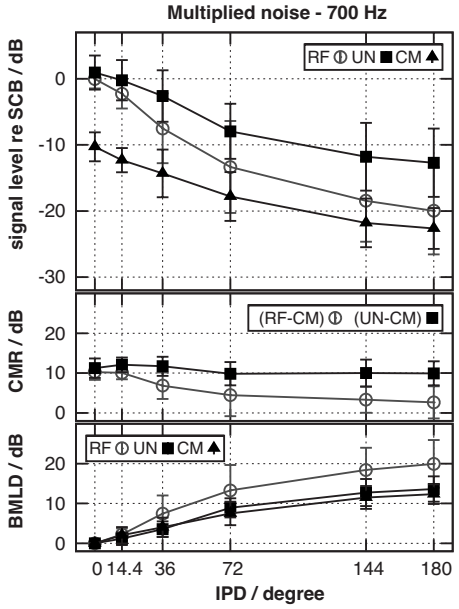


FIG. 2. Results for the multiplied noise masker. The upper panel shows mean detection thresholds averaged over all listeners for the reference (RF, circles), uncorrelated (UN, squares), and comodulated (CM, triangles) conditions. Thresholds are plotted relative to the level of the masker centered at the signal frequency (SCB). The middle panel shows the average CMR(RF-CM) (circles) and CMR(UN-CM) (squares). In the lower panel, the masking release (BMLD) relative to the diotic condition is shown for RF (circles), UN (squares), and CM (triangles). Error bars indicate ±1 standard deviation.

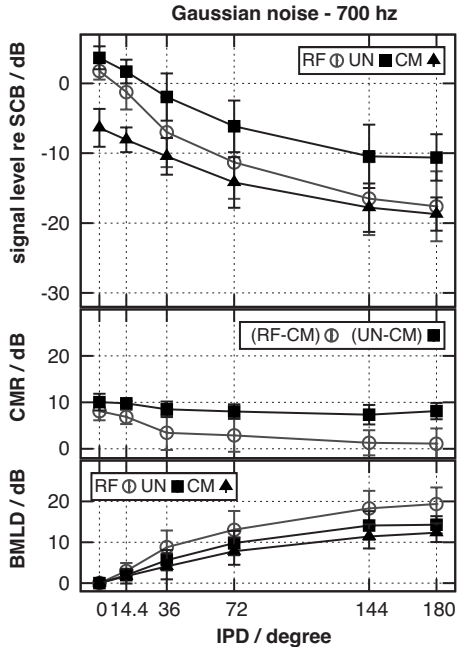


FIG. 3. As Fig. 2 but for the Gaussian noise masker.

listeners showed no or only minor differences in the BMLD between single-band and multi-band conditions.

The diotic CMR(UN-CM) varied from 7 dB (listener AH) to 15 dB (listener TK). For the majority of the listeners, the CMR(UN-CM) in dichotic listening conditions was approximately constant with IPD. A few listeners showed a slight decrease (BE, JW, and MK) or increase (TK and RE) of the CMR(UN-CM) with increasing IPD. For most listeners, the CMR (RF-CM) was very similar to the CMR(UN-CM) in the diotic condition but differed from the CMR(UN-CM) in the dichotic conditions. In contrast to CMR(UN-CM), the CMR(RF-CM) decreased for most listeners with increasing IPD. Only listeners AH and MK showed a similar CMR(UN-CM) and CMR(RF-CM).

Figure 2 shows mean results for the data shown in Fig. 1 with interindividual standard deviation. In the upper panel, thresholds are plotted as in Fig. 1. In the middle panel, the CMR is shown for each value of the IPD. Circles and squares indicate CMR(RF-CM) and CMR(UN-CM), respectively. The lower panel shows the BMLD, i.e., the difference in threshold for each dichotic condition (IPD $\neq 0$) relative to the threshold for the corresponding diotic condition (IPD=0). As in the upper panel, circles, squares, and triangles indicate BMLDs for the reference, uncorrelated, and comodulated conditions, respectively.

The average thresholds show a monotonic decrease with increasing IPD. The magnitude of the CMR in the diotic condition was similar for the two definitions of CMR: The diotic CMR(RF-CM) was about 10 dB (middle panel, circles) and the diotic CMR(UN-CM) was about 11 dB (middle panel, triangles). By definition, the BMLD (lower panel) was zero for no IPD. The improvement in thresholds with increasing IPD is reflected in a monotonic increase in the BMLD. The maximum BMLD occurred for an antiphasic signal and was 20 dB for the reference condition and about 13 dB for the uncorrelated and comodulated conditions.

The CMR(UN-CM) was almost constant for the different values of the IPD (deviation of less than 1 dB from the mean value). In contrast, CMR(RF-CM) decreased monotonically as the IPD increased to a minimum value of about 3 dB.

The standard deviation of the thresholds (upper panel of Fig. 2) increased with increasing IPD. In contrast, similar standard deviations of CMR(UN-CM) (middle panel of Fig. 2) and the BMLD were found for all IPDs for the uncorrelated and comodulated conditions (lower panels of Fig. 2). This reflects the fact that the individual masked thresholds varied (see Fig. 1), but the masking releases were very similar for all listeners. The standard deviations were larger for the thresholds in the reference condition than for thresholds in the uncorrelated and comodulated conditions. CMR(RF-CM) also showed larger variability than CMR(UN-CM) and the variability increased when an IPD was introduced.

Figure 3 shows mean results for the Gaussian noise masker, plotted in the same manner as in Fig. 2. Relative to the thresholds for the multiplied noise masker, all thresholds were elevated by about 3 dB. As for the multiplied noise masker, thresholds decreased monotonically with increasing IPD. The CMR(UN-CM) hardly changed (deviation of less than 1 dB from mean value) with increasing IPD while CMR(RF-CM) decreased monotonically with increasing IPD. The maximum BMLD was almost identical to that for the multiplied noise masker (Fig. 2).

For both masker types, the constant CMR(UN-CM) indicates that the auditory system can use comodulation to achieve a masking release in dichotic listening conditions to the same extent as in diotic listening conditions, i.e., the

masking released add in decibels. On the other hand, the reduction in CMR(RF-CM), where a single-band condition is compared to a multi-band condition, indicates that, for an antiphasic signal, there is little extra benefit from the co-modulated masker bands (circles in middle panel of Fig. 3).

IV. EXPERIMENT 2: CMR AND BMLD FOR HIGHER AND LOWER FREQUENCIES

A. Rationale

To investigate if the additivity of CMR and BMLD in decibels also holds for other frequencies, signal frequencies both below and above 700 Hz were used. For lower frequencies, only a small CMR is to be expected (Schooneveldt and Moore, 1987), while a large BMLD should occur (van de Par and Kohlrausch, 1999). On the other hand a large CMR is to be expected at a higher signal frequency (Schooneveldt and Moore, 1987), while the BMLD is reduced for frequencies above about 1500 Hz due to a gradual loss of fine-structure information (van de Par and Kohlrausch, 1997). Thus, data at various center frequencies provide insight into the combined effect of comodulation and interaural disparities with different magnitudes of the single effects CMR and BMLD.

B. Methods

The center frequencies of the masker bands were chosen to have the same ratios of signal frequency and masker band center frequency as those used in the first experiment. The FBs were located at 85, 115, 285, and 315 Hz for the 200-Hz signal and at 1285, 1715, 4285, and 4715 Hz for the 3000-Hz signal. The signal IPD was 0°, 72°, or 180°. The stimulus was presented using Sennheiser HD 650 headphones.

C. Results

Only mean data are shown since the inter-subject variablility was similar to that for the data of Experiment I. Figure 4 shows mean data for the signal frequency of 200 Hz. The left and right panels show results for multiplied and Gaussian noise maskers, respectively. Thresholds are plotted in the upper row. The middle row shows CMR(RF-CM) and CMR(UN-CM) and the lower row shows the BMLD for the reference, uncorrelated, and comodulated conditions using the same symbols as in Fig. 2. Compared to the data obtained at 700 Hz, the CMR was smaller while the BMLD was larger for the multi-band conditions (uncorrelated and comodulated) and slightly smaller for the single-band (reference) condition.

For the multiplied noise masker (left panels), the diotic CMR(RF-CM) was about 5 dB and the diotic CMR(UN-CM) was about 8 dB. The thresholds decreased monotonically with increasing IPD. The CMR(RF-CM) and the CMR(UN-CM) hardly varied with the IPD (deviation of less than 1 dB from mean value) and the maximum BMLD was about 17 dB and for all three conditions.

For the Gaussian noise masker (right panels), the diotic CMR(RF-CM) was about 0 dB, while CMR(UN-CM) was about 4 dB. The decrease in thresholds with increasing IPD

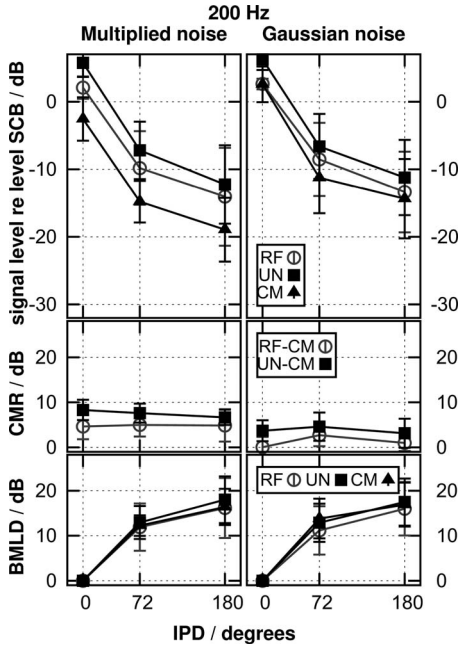


FIG. 4. Results for a signal frequency of 200 Hz with multiplied and Gaussian noise maskers (left and right panels, respectively). Otherwise, as Fig. 2.

was very similar to that for the multiplied noise masker. The CMR(RF-CM) and CMR(UN-CM) varied little with the IPD (deviation of less than 2 dB and less than 1 dB from mean value, respectively). The maximum BMLD was about 17 dB.

The standard deviations of the thresholds (upper panels) increased as the IPD increased. The standard deviations were largest for the reference condition. In contrast, the magnitude of CMR (middle panels) showed only small standard deviations across listeners. The same trend was observed for the data for the 700-Hz signal. The variability of the BMLD (lower panels) was larger than the variability of the CMR and was largest for the BMLD of the reference condition, reflecting larger individual differences in binaural performance than in the performance for processing across-frequency cues.

Thresholds for the Gaussian noise masker were slightly higher than for the multiplied noise masker and the CMR was slightly smaller, while the BMLD was very similar for the two masker types. The CMR was more affected by the type of noise for the center frequency of 200 Hz than for the center frequency of 700 Hz.

Figure 5 shows results for the signal frequency of 3000 Hz for the multiplied (left) and Gaussian (right) noise maskers. The CMR was larger than for the 200-Hz signal and the BMLD was smaller than at 200 and 700 Hz.

For the multiplied noise masker, the diotic CMR(RF-CM) was 11 dB and the diotic CMR(UN-CM) was 13 dB. As for the other signal frequencies, thresholds decreased monotonically with increasing IPD. The CMR(RF-CM) and CMR(UN-CM) hardly varied with IPD. The maximum BMLD was about 4 dB for all three conditions.

As for the other signal frequencies, thresholds for the Gaussian noise masker were slightly higher than thresholds for the multiplied noise masker and the CMR was smaller by about 5 dB. The diotic CMR(RF-CM) was about 6 dB and

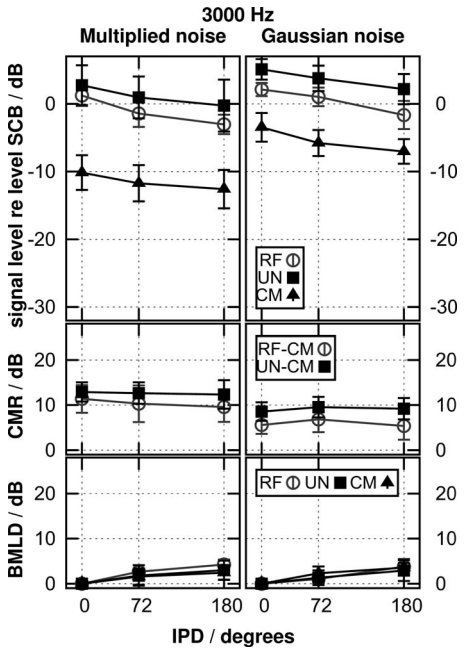


FIG. 5. As Fig. 4, but for a signal frequency of 3000 Hz.

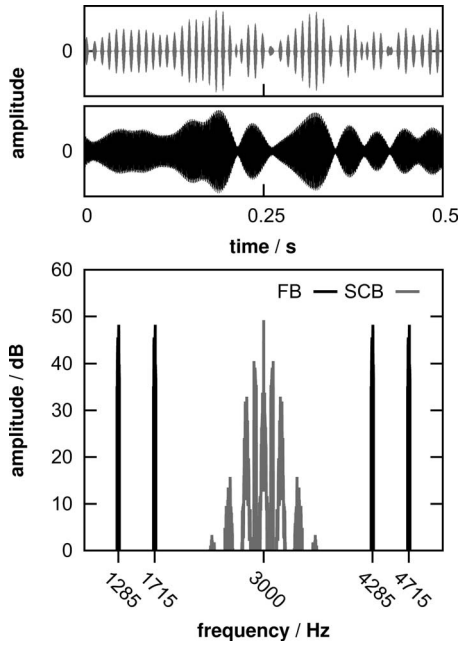


FIG. 6. Time signal and spectrum of a Gaussian noise masker sample. Upper panel: Time signals of the SCB (gray) and the FB (black) centered at 4285 Hz. Lower panel: Spectrum of transposed masker band centered at the signal frequency (SCB) together with additional FBs. The SCB shows additional sidebands as a result of the transposition, as described by van de Par and Kohlrausch (1997). The levels of the added FBs are well above the levels of the sidebands generated by the transposition of the SCB.

the diotic CMR(UN-CM) was 9 dB. The CMR hardly varied with IPD. The maximum BMLD was similar to that for the multiplied noise masker (4 dB).

Thresholds for the Gaussian noise masker were slightly higher and the CMR was slightly smaller than for the multiplied noise masker. The BMLD was similar for the two noise types and across masker conditions.

The data for signal frequencies of 200 and 700 Hz show the ability to process across-frequency cues in dichotic listening conditions where the BMLD is larger than the CMR. The data for the signal frequency of 3000 Hz also show a constant CMR, independent of the IPD. However, the small magnitude of the BMLD at this center frequency does not allow a clear interpretation of the nature of the combined effect.

V. EXPERIMENT 3: TRANSPOSED STIMULUS

A. Rationale

The maximum BMLD for the signal frequency of 3000 Hz was considerably smaller than the CMR at this frequency. To further investigate how CMR and BMLD combine at high frequencies, a transposed stimulus was used. The transposed stimulus has been shown to increase the BMLD at high signal frequencies (van de Par and Kohlrausch, 1997). In the present study, the SCB with the signal was transposed from 100 to 3000 Hz. This transposition introduced fluctuations at multiples of 100 Hz into the temporal envelope of the SCB. Thus, even in the comodulated condition, the transposed SCB had a slightly different envelope than the FBs. The data of Eddins and Wright (1994) suggest that the auditory system is able to make an across-frequency envelope comparison at different envelope rates simultaneously. In the case where, for example, the modulation is coherent at only one modulation frequency, only the effect for this modulation frequency should be observed. This implies that, for the transposed condition, a CMR for the low rate should still be observed even if the FBs and the SCB are different with respect to the high envelope frequency components but are identical in the low envelope frequency components.

B. Methods

The SCB was generated at a center frequency of 100 Hz and transposed to a center frequency of 3000 Hz. In the target interval, a 100-Hz sinusoidal signal was added to the masker band prior to transposition. The transposition was done by half-wave rectification and lowpass filtering (second order, 500-Hz cut-off frequency) of the low-frequency waveform and subsequent multiplication with a 3000-Hz carrier.

Multiplied or Gaussian noise flanking masker bands were added at the same frequencies as in the second experiment. In order to avoid spectral overlap of the sidebands introduced by the transposition, the FBs were not transposed from low frequencies to high frequencies but were generated as in the experiment for the non-transposed 3000-Hz center frequency. Figure 6 shows the time signals (upper panels) and the spectrum (lower panel) of the SCB (gray) and the FBs (black). There are sidebands around the SCB as a result of the transposition. The FBs were centered at 1285, 1715, 4285, and 4715 Hz. The stimulus was presented using Sennheiser HD 650 headphones.

C. Results

Figure 7 shows mean results for the transposed stimulus in the same format as Fig. 5. The results were comparable to

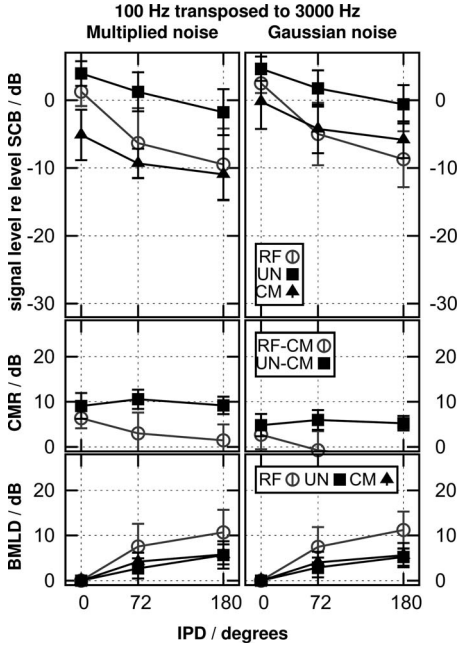


FIG. 7. As Fig. 5, but for the transposed stimulus.

those obtained for the non-transposed 3000-Hz signal. The CMR was slightly smaller and the BMLD was larger for the transposed stimulus than for the non-transposed stimulus.

For the multiplied noise masker, the diotic CMR(UN-CM) was about 10 dB and varied little with IPD. The diotic CMR(RF-CM) was about 6 dB and decreased with increasing IPD to 1 dB for an antiphasic signal. The maximum BMLD was larger than for the 3000-Hz condition where the SCB was not transposed (see Sec. IV). It was 11 dB for the reference condition and about 6 dB for the uncorrelated and comodulated conditions.

The diotic CMR(UN-CM) for the Gaussian noise masker was about 5 dB and did not vary with IPD. The diotic CMR(RF-CM) was about 3 dB and reduced with increasing IPD to a negative value for the antiphasic signal. The BMLDs were about the same as for the multiplied noise masker.

In contrast to the data where the SCB was not transposed (Fig. 5), the standard deviations of the thresholds (upper row of Fig. 7) indicate large individual differences in masked thresholds. But, as for the data where the SCB was not transposed, the CMR(UN-CM) showed only small variation across listeners. The same holds true for the BMLDs for the uncorrelated and comodulated conditions for both masker types. The standard deviations of the CMR(RF-CM) and of the BMLD for the reference condition increased with increasing IPD.

Thresholds for the Gaussian noise masker were slightly higher than thresholds for the multiplied noise masker. The CMR(RF-CM) was smaller for the Gaussian noise masker than for the multiplied noise masker and was even negative for higher values of the IPD. For both masker types, the CMR(UN-CM) did not vary with IPD. In contrast, the BMLDs obtained for the reference, uncorrelated, and comodulated conditions were almost identical for the two masker types.

VI. DISCUSSION

Epp and Verhey (2009) showed that CMR(UN-CM) and BMLD are additive in decibels at a signal frequency of 700 Hz. Their data indicated that the ability to use comodulation cues across frequency was not affected by the introduction of binaural cues and vice versa. The results of the present study support this hypothesis for the 700-Hz frequency for a different set of subjects. In addition, the data show that CMR(UN-CM) and BMLD are additive for signal frequencies of 200 and 3000 Hz. The additivity does not seem to be restricted to a particular signal frequency. Thus, the cues comodulation and IPD appear to be processed independently in the auditory system. Within the same set of subjects, the additivity is not found for CMR(RF-CM) and BMLD.

A. Comparison with previous studies focusing on either CMR or BMLD

In general, the results on CMR in the diotic condition (i.e., with no IPD) and those on the BMLD with an antiphasic signal (IPD of 180°) are in good agreement with data found in the literature (Hall *et al.*, 1990; Schooneveldt and Moore, 1989; van de Par and Kohlrausch, 1999).

Hall *et al.* (1990) used 20-Hz-wide multiplied noise masker bands with the same center frequencies as used in the first experiment of the present study and a signal frequency of 700 Hz. They reported a CMR(RF-CM) of about 11 dB. This is in good agreement with the CMR(RF-CM) found in the present study (10 dB). Data from Schooneveldt and Moore (1989) with one FB only indicated an increase in diotic CMR(UN-CM) from 5 to 8 dB as the center frequency was increased from 250 to 4000 Hz. A similar effect of center frequency was found in the present study: for a center frequency of 200 Hz, CMR(UN-CM) was 7 dB for the multiplied noise masker and 3 dB for the Gaussian noise masker. For a signal frequency of 3000 Hz, the CMR(UN-CM) was about 6 dB larger than for the 200-Hz signal. The differences in the magnitude of CMR between the present study and the study of Schooneveldt and Moore (1989) are presumably due to the higher number of FBs used in the present study.

In agreement with Epp and Verhey (2009), the present data indicate that the thresholds for Gaussian noise maskers are higher than corresponding thresholds for multiplied noise maskers and that the CMR tends to be slightly smaller for Gaussian noise than for multiplied noise. This could be explained in terms of "listening in the valleys" (Buus, 1985) since Gaussian noise has less envelope amplitude values close to zero, at which the instantaneous signal-to-noise ratio is high and might be used to improve detectability.

In the present study, the BMLD for the reference condition and a signal IPD of 180°, was about 12 dB larger for center frequencies of 200 and 700 Hz than for the center frequency of 3000 Hz. A similar trend was observed by van de Par and Kohlrausch (1999). For a diotic 25-Hz-wide Gaussian noise masker and an antiphasic signal, they found BMLDs of about 25 dB at 500 Hz, 23 dB at 250 Hz, and 8 dB at 4000 Hz.[1] In the present study, the BMLD for the

single-band condition with a 3000-Hz signal was increased by about 7 dB when a transposed stimulus was used. This increase agrees with the results of van de Par and Kohlrausch (1997), who used a stimulus centered at 125 Hz with a masker bandwidth of 25 Hz which was transposed to 4000 Hz.

For each signal frequency, the maximum BMLDs were the same for the two multi-band conditions (uncorrelated and comodulated conditions) and did not differ between the two masker types. The maximum BMLDs for the multi-band conditions were smaller than for the single-band masker (reference condition). A similar effect was found in previous BMLD studies where a reduced BMLD was found when the masker bandwidth was increased. It was hypothesized by van de Par and Kohlrausch (1999) that this reduction might be due to the hampered ability to use off-frequency filters in such spectrally broad conditions compared to narrowband conditions. In terms of this deterioration, the addition of FBs might have the same effect as the broadening of the masker.

For a center frequency of 1000 Hz, van de Par and Kohlrausch (1999) found a reduction in the BMLD from 25 to 16 dB when the masker bandwidth was increased from 25 to 1000 Hz. At 4000 Hz, the reduction was only 2 dB. In line with their data, the difference in BMLD between the single-band condition and the multi-band conditions in the present study was about 6 dB for 700 Hz and less than 2 dB at 3000 Hz. Note that the difference was also 6 dB at a signal frequency of 3000 Hz when transposed signals were used.

For the 200-Hz signal frequency, hardly any difference in the BMLD was observed between the single-band and multi-band conditions, whereas for a comparable signal frequency (250 Hz), van de Par and Kohlrausch (1999) found a substantial decrease in the BMLD as the bandwidth increased. The difference between the studies may reflect individual differences. The individual data for the signal frequency of 700 Hz (Fig. 1) show that at least two subjects (MK and AH) had a similar BMLD for the multi-band and single-band conditions, although the majority of the listeners showed a larger BMLD for the single-band (reference) condition.

In general, the highest variability in the thresholds and the BMLDs was found for the narrowband reference condition. This finding is in agreement with results of Buss *et al.* (2007). On the basis of their results, Buss *et al.* (2007) argued that the large individual differences in binaural performance for narrowband maskers may indicate that there are good and poor binaural listeners in experiments using narrowband noise as masker.

Note that the BMLD, in contrast to the CMR, does not show any dependence on the statistics of the masker. The difference in CMR between multiplied and Gaussian maskers might be related to the smaller modulation depth of Gaussian noise compared to multiplied noise. In terms of a listening in the valleys approach (Buus, 1985), a smaller modulation depth leads to a lower signal-to-noise ratio in the valleys and consequently to increased thresholds. Such an effect of modulation depth had been observed by Verhey *et al.* (1999) in a bandwidening type of CMR experiment and by Eddins (2001) for a FB type of experiment using low-noise noise.

For the BMLD, the data may be understood in terms of an "equalization-cancellation mechanism" (Durlach, 1963). Cancellation of a diotic masker is independent of the actual statistical properties of the masker under the assumption that the cancellation is performed in an optimal way or that the error of the cancellation is negligible.

B. Comparison to previous data on the combined effect of interaural differences and comodulation

There are only a few studies on CMR with dichotic signals. These studies differ in their results and their interpretation of the data. Using a FB paradigm, Hall *et al.* (1988) found a decrease in CMR(RF-CM) in dichotic conditions compared to diotic conditions: On average, CMR(RF-CM) was 7 dB in the diotic condition and 2 dB in the dichotic condition. While half of the subjects had a small CMR(RF-CM), the other listeners did not benefit from the additional comodulated masker band. They concluded that there are large individual differences in benefit due to an added comodulated masker band in dichotic listening conditions. Cohen and Schubert (1991) found a diotic CMR(RF-CM) of about 6 dB and a negligible CMR(RF-CM) when the signal was presented with an IPD of 180°. The CMR(UN-CM) decreased from 9 to 4 dB when the signal phase was changed from 0 to 180°. Thus a residual CMR was observed when the uncorrelated condition was used as a reference. On the basis of this result, Cohen and Schubert (1991) concluded that CMR and BMLD are additive to some extent.

In a recent study, Hall *et al.* (2006) found a large influence of the definition of CMR on the difference between the CMR for a diotic and a dichotic signal. When the signal IPD was changed from 0 to 180°, the CMR(RF-CM) decreased from 12 to less than 2 dB while the CMR(UN-CM) decreased from 12 to about 5 dB. Compared to the present study, the diotic CMR(UN-CM) is larger in the study of Hall *et al.* (2006) while the dichotic CMR(UN-CM) is smaller. This might be due to differences in the experimental parameters e.g., the spectral distance between the FBs and the SCB. Due to this smaller spectral distance, it is more likely that within-channel CMR contributed to the masking release, leading to a larger CMR in the diotic condition. Hall *et al.* (2006) used experimentally derived psychometric functions in connection with a signal-detection-theory approach to investigate the mechanism of the thresholds obtained for a comodulated masker and dichotic signal presentation. They showed that their data could not be modeled using addition of d' within an integration model (Green and Swets, 1988). They concluded that the combined effect of comodulation and binaural cues is larger than would be expected from a simple addition of the d' values.

A similar influence of the definition of CMR was observed in the present study for a comparable signal frequency (700 Hz). The CMR(RF-CM) decreased from about 8 to 10 dB (depending on the masker type) for the diotic signal to 1–3 dB for the dichotic signal. A similar decrease

was found for the transposed stimuli. In contrast, CMR(UN-CM) showed a reduction in the masking release of less than 1 dB. The reduced CMR(RF-CM) in the dichotic condition found by Hall *et al.* (2006) is in qualitative agreement with the reduced CMR(RF-CM) found in the present study for all dichotic conditions. But there is one clear difference: while Hall *et al.* (2006) observed a decrease in CMR(UN-CM) when changing the interaural signal phase from 0 to 180°, the results of the present study indicate that the CMR(UN-CM) is independent of the IPD, i.e., the two masking releases add. This qualitative difference may be due to the differences in the experimental parameters. It is likely that the CMR found by Hall *et al.* (2006) was mainly due to within-channel processes since they used a small spectral distance between the SCB and the nearest FB.[2] In contrast, the present study used a three times larger spectral distance. McFadden (1986) and Schooneveldt and Moore (1987) suggested that a large part of CMR in a FB experiment with close proximity of the on- and off-frequency masker components may be due to within-channel cues rather than reflecting an across-channel effect. In line with that hypothesis, Piechowiak *et al.* (2007) showed that, for the spectral distance (relative to the signal frequency) comparable to the one used in Hall *et al.* (2006), most of the CMR(UN-CM) can be predicted by a within-channel model. For the spectral distance used in the present study, the model predicted a negligible CMR. This indicates that the CMR found in the present study was mainly due to across-channel processes. Thus, as CMR(UN-CM) was found to be constant with the IPD for the spectral distance used in the present study, CMR(UN-CM) may only be independent of the IPD if it mainly results from across-channel processes. A hypothesis that can be derived from this result is that, in conditions where the diotic CMR is due to a combination of within- and across-channel contributions, the dichotic CMR only reflects the part of CMR that is due to across-channel processes.

C. Implications for the underlying mechanism

The influence of the definition of the CMR on the effect of IPD complicates the interpretation of the nature of the combination of monaural across-channel cues and binaural cues. Thus, before discussing the possible underlying mechanism it is important to understand the effect of the definition of CMR on the results. In this context it is interesting to reconsider the difference in BMLD for the different masking conditions.

The data of Hall *et al.* (1988) showed that the BMLD with no flanker present had an average value of about 22 dB, whereas the BMLD for the comodulated condition was only 17 dB. A similar reduction in the BMLD was found by Cohen and Schubert (1991), Hall *et al.* (2006), and in the present study.

The main difference between the reference and comodulated conditions, apart from the comodulation, is the spectrum of the masker. As mentioned before, an increase in the number of spectral components might have a similar effect on the BMLD as an increase in bandwidth which decreases the BMLD. Thus the difference in the BMLD for the spectrally broad (uncorrelated and comodulated) and spectrally narrow conditions (reference) might simply reflect the hampered ability to use off-frequency filters as proposed as an explanation for the bandwidth dependence of the BMLD (van de Par and Kohlrausch, 1999). Hence, the reduction in CMR(RF-CM) might be the result of two effects: A reduction in threshold due to comodulation and an increase in threshold due to a change in the masker spectrum. The results for the uncorrelated condition of the present study are consistent with the hypothesis that the reduction in the dichotic compared to the diotic CMR(RF-CM) reflects the reduced ability of the binaural system to use off-frequency information in the comodulated condition compared to the reference condition rather than the reduced ability of the auditory system to use across-frequency information.

This hypothesis is supported by the data of Hall *et al.* (2006) using different interaural correlation of the masker. Reducing the interaural correlation to 0.95 abolishes the difference in the BMLD for the reference and the uncorrelated conditions since for this reduced interaural correlation beneficial across-channel processes can no longer be used by the auditory system (van de Par and Kohlrausch, 1999). As a consequence, the same dichotic CMR is obtained with the two definitions of CMR. The smaller dichotic CMR(UN-CM) reported by Hall *et al.* (2006) is presumably a consequence of the large contribution of within-channel cues to the CMR in their study (see above). The results of the present study indicate that the single-band (reference, RF) condition is not an appropriate reference to study the combined effect of monaural and binaural cues due to (i) large interindividual difference and (ii) effects due to differences in spectra changing the ability to use off-frequency information in dichotic conditions.

The present data for CMR(UN-CM) indicate that the two masking releases are additive in decibels, i.e., the overall masking release is the sum of the CMR(UN-CM) and the BMLD in decibels at each value of the IPD. The additivity of across-channel CMR and BMLD might provide insights into the topographic organization of the processing stages involved in CMR and BMLD processing: Epp and Verhey (2009) showed that a conceptual model based on serial processing stages was able to account for the data from a combined CMR(UN-CM) and BMLD experiment at a signal frequency of 700 Hz. The data of Hall *et al.* (1988), Cohen and Schubert (1991), and Hall *et al.* (2006) show a combined effect less than a summation, which is presumably due to the influence of within-channel cues and the comparison of conditions with different spectra, i.e., CMR(RF-CM) in dichotic CMR paradigms. This difference in the ability to combine CMR and BMLD with small and large spectral distances of the FBs from the signal frequency may serve to disentangle within-channel and across-channel contributions in CMR paradigms.

The data for the transposed stimuli are in line with the hypothesis that comodulation and IPD cues are processed independently in the auditory system. The results show a similar magnitude of the CMR as for the data of the nontransposed 3000-Hz stimulus. The similarity in the CMR suggests that the processing stage of CMR is unaffected by

the transposition, presumably since the across-frequency envelope cues are preserved. In contrast, the BMLD is increased for the transposed stimuli since this procedure introduces interaural time delay cues that would normally only be available at low signal frequencies.

The invariance of CMR(UN-CM) with the IPD at various magnitudes of the CMR(UN-CM) and the BMLD has two implications for the underlying mechanisms. First, the performance of the processing stage which uses either comodulation or IPD is not affected by the other cue, i.e., these two cues seem to be processed independently in the auditory system. Second, the additivity of CMR(UN-CM) and BMLD in decibels can be interpreted as a progressive improvement of the internal representation of the masked signal along the ascending auditory pathway. Such an improvement could be realized as a serial alignment of the underlying processing stages.

VII. SUMMARY AND CONCLUSIONS

We investigated the ability of the auditory system to benefit from processing of across-frequency and across-ear cues simultaneously using CMR experiments with FBs and various IPDs of the signal. The results show the following.

(i) CMR(RF-CM) and CMR(UN-CM) were similar in a diotic condition, but differed in dichotic conditions. While CMR(RF-CM) decreased with the introduction of an IPD, CMR(UN-CM) was almost unaffected by an interaural phase difference of the signal.

(ii) The decrease in CMR(RF-CM) in dichotic listening conditions may reflect the reduced ability to use off-frequency filters to process interaural disparities (which was hypothesized to explain the effect of bandwidth on the BMLD) rather than a reduced ability to process comodulation in a dichotic listening condition. Thus, the reference (RF) condition of the CMR FB paradigm might be a problematic reference if the influence of binaural cues is investigated since the masker spectrum for the reference condition is different from that for the other two conditions (UN and CM). In addition, CMR(RF-CM) strongly depends on the IPD, and reference thresholds show a large variability across subjects. This also hampers the interpretation of the combination of monaural across-channel and binaural cues.

(iii) The comparison with the uncorrelated condition does not suffer from the interfering effect of the width of the spectrum covered by masker components on the magnitude of the BMLD. For this comparison, a summation of the benefit due to comodulation [CMR(UN-CM)] and the benefit due to an IPD (BMLD) was found. The addition was also found for a transposed stimulus. This indicates that the uncorrelated condition is a less problematic reference for quantification of the masking release due to comodulation in dichotic listening conditions. CMR(UN-CM) is not dependent on the IPD and shows only small variability across subjects.

(iv) The additivity of across-channel CMR and BMLD holds true for multiplied noise as well as for Gaussian noise, i.e., the additivity of CMR and BMLD is independent of the envelope amplitude distribution of the masking noise used in the present study.

(v) The same CMR(UN-CM) for all IPDs suggests independent processing of CMR and BMLD and supports the hypothesis of Schooneveldt and Moore (1989) and Epp and Verhey (2009) of a serial arrangement of the processing stages underlying CMR and BMLD.

ACKNOWLEDGMENTS

We would like to thank the associate editor Brian Moore, Joe Hall, and one anonymous reviewer for many helpful comments on a previous version of this manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (International Graduate School for "Neurosensory Science and Systems" GRK 591 and SFB TR31).

[1]Note that, for all signal frequencies, the magnitude of BMLD reported in the study of van de Par and Kohlrausch (1999) is larger than in the present study. This may be partly due to individual differences. Only three subjects participated in their study, among them the two authors who are certainly both highly trained in binaural listening tasks. An additional problem for the comparison of broadband data with data of a multi-band paradigm is the difference in spectral content of the masker. The spectral notches in a flanking paradigm lead to differences in the modulation spectrum compared to a masker with a continuous spectrum and the same minimum and maximum frequencies. The comparison with data from the literature using a similar spectral range was included here since it was the most comparable data set to the masking condition used in the multi-band conditions.

[2]In one experiment, Hall *et al.* (2006) also used a larger spectral distance between the SCB and the FBs. However, for this spectral distance they did not measure thresholds for the uncorrelated condition.

Buus, S. (**1985**). "Release from masking caused by envelope fluctuations," J. Acoust. Soc. Am. **78**, 1958–1965.

Buss, E., Hall, J. W., and Grose, J. H. (**2007**). "Individual differences in the masking level difference with a narrowband masker at 500 or 2000 Hz," J. Acoust. Soc. Am. **121**, 411–419.

Cohen, M. (**1991**). "Comodulation masking release over a three octave range," J. Acoust. Soc. Am. **90**, 1381–1384.

Cohen, M. F., and Schubert, E. D. (**1991**). "Comodulation masking release and the masking level-difference," J. Acoust. Soc. Am. **89**, 3007–3008.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.

Eddins, D. A. (**2001**). "Monaural masking release in random-phase and low-noise noise," J. Acoust. Soc. Am. **109**, 1538–1549.

Eddins, D., and Wright, B. A. (**1994**). "Comodulation masking release for single and multiple rates of envelope fluctuation," J. Acoust. Soc. Am. **96**, 3432–3442.

Epp, B., and Verhey, J. L. (**2009**). "Superposition of masking releases," J. Comput. Neurosci. **26**, 393–407.

Ernst, S., and Verhey, J. L. (**2006**). "Role of suppression and retro-cochlear processes in comodulation masking release," J. Acoust. Soc. Am. **120**, 3843–3852.

Green, D. M., and Swets, J. A. (**1988**). *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966) (Reprinted by Peninsula, Los Altos, CA, **1988**).

Hall, J., Buss, E., and Grose, J. (**2006**). "Binaural comodulation masking release: Effects of masker interaural correlation," J. Acoust. Soc. Am. **120**, 3878–3888.

Hall, J. W., Cokely, J.-A., and Grose, J. H. (**1988**). "Combined monaural and binaural masking release," J. Acoust. Soc. Am. **83**, 1839–1845.

Hall, J. W., Grose, J. H., and Haggard, M. P. (**1990**). "Effects of flanking band proximity, number, and modulation pattern on comodulation masking release," J. Acoust. Soc. Am. **87**, 269–283.

Hall, J. W., Haggard, M. P., and Fernandes, M. A. (**1984**). "Detection in noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am. **76**, 50–56.
Hirsh, I. J. (**1948**). "The influence of interaural phase on interaural summation and inhibition," J. Acoust. Soc. Am. **20**, 536–544.
Jeffress, L. A., Blodgett, H. C., Sandel, T. T., and Wood, C. L., III (**1956**). "Masking of tonal signals," J. Acoust. Soc. Am. **28**, 416–426.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Licklider, J. C. R. (**1948**). "The influence of interaural phase relations upon the masking of speech by white noise," J. Acoust. Soc. Am. **20**, 150–159.
McFadden, D. (**1986**). "Comodulation masking release: Effects of varying the level, duration, and time delay of the cue band," J. Acoust. Soc. Am. **80**, 1658–1667.
Moore, B. C. J., Hall, J. W., Grose, J. H., and Schooneveldt, G. P. (**1990**). "Some factors affecting the magnitude of comodulation masking release," J. Acoust. Soc. Am. **88**, 1694–1702.
Nelken, I., Rotman, Y., and Yosef, O. (**1999**). "Responses of auditory cortex neurons to structural features of natural sounds," Nature (London) **397**, 154–157.
Piechowiak, T., Ewert, S. D., and Dau, T. (**2007**). "Modeling comodulation masking release using an equalization-cancellation mechanism," J. Acoust. Soc. Am. **121**, 2111–2126.
Schooneveldt, G. P., and Moore, B. C. J. (**1987**). "Comodulation masking release (CMR): Effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band," J. Acoust. Soc. Am. **82**, 1944–1956.
Schooneveldt, G. P., and Moore, B. C. J. (**1989**). "Comodulation masking release for various monaural and binaural combinations of the signal, on-frequency, and flanking-bands," J. Acoust. Soc. Am. **85**, 262–272.
van de Par, S., and Kohlrausch, A. (**1997**). "A new approach to comparing binaural masking level differences at low and high frequencies," J. Acoust. Soc. Am. **101**, 1671–1680.
van de Par, S., and Kohlrausch, A. (**1999**). "Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters," J. Acoust. Soc. Am. **106**, 1940–1947.
Verhey, J. L., Dau, T., and Kollmeier, B. (**1999**). "Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation filterband model," J. Acoust. Soc. Am. **106**, 2733–2745.
Verhey, J., Pressnitzer, D., and Winter, I. (**2003**). "The psychophysics and physiology of comodulation masking release," Exp. Brain Res. **153**, 405–417.
Verhey, J., Rennies, J., and Ernst, S. (**2007**). "Influence of envelope distributions on signal detection," Acta. Acust. Acust. **93**, 115–121.
Zurek, P., and Durlach, N. (**1987**). "Masker-bandwidth dependence in homophasic and antiphasic tone detection," J. Acoust. Soc. Am. **81**, 459–464.

Investigating possible mechanisms behind the effect of threshold fine structure on amplitude modulation perception[a)]

Stephan J. Heise,[b)] Manfred Mauermann, and Jesko L. Verhey
Institut für Physik, Universität Oldenburg, D-26111 Oldenburg, Germany

(Received 10 February 2009; revised 18 June 2009; accepted 17 August 2009)

Detection thresholds for sinusoidal amplitude modulation at low levels are higher (worse) when the carrier of the signal falls in a region of high pure-tone sensitivity (a minimum of the fine structure of the threshold in quiet) than when it falls at a fine-structure maximum. This study explores possible mechanisms behind this phenomenon by measuring modulation detection thresholds as a function of modulation frequency (experiment 1) and of carrier level for tonal carriers (experiment 2) and for 32-Hz wide noise carriers (experiment 3). The carriers could either fall at a fine-structure minimum, a fine-structure maximum, or in a region without fine structure. Modulation frequencies varied between 8 Hz and one fine-structure cycle, and carrier levels varied between 7.5 and 37.5 dB sensation levels. A large part of the results can be explained by assuming a reduction in effective modulation depth by spontaneous otoacoustic emissions—or more generally cochlear resonances—that synchronize to the carrier at fine-structure minima. Beating between cochlear resonances and the stimulus ("monaural diplacusis") may hamper the detection task, but this cannot account for the whole effect. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3224731]

PACS number(s): 43.66.Mk, 43.66.Cb, 43.64.Jb [MW] Pages: 2490–2500

I. INTRODUCTION

The perception of amplitude modulation (AM) at high and moderate levels has been the subject of many studies in auditory research due to its relevance in everyday hearing. Amplitude modulations are believed to be detected on the basis of temporal fluctuations in the envelope, at least for modulation frequencies below about 100 Hz (e.g., Yost and Sheft, 1997). At moderate and high levels this leads to modulation detection thresholds (MDTs) of sinusoidally amplitude modulated tones being practically independent of carrier frequency and of modulation frequency (Kohlrausch *et al.*, 2000; Moore and Glasberg, 2001). At low levels, however, the spectral characteristics of the stimulus may have an influence on AM detection in regions where the threshold in quiet shows what is known as threshold fine structure.

Threshold fine structure—sometimes also referred to as microstructure—are quasi-periodic oscillations in the threshold in quiet, which are found in normal-hearing listeners (Elliott, 1958). The amplitude of these oscillations may be as high as 15 dB and the typical periodicity is around 1/10 octave. It has been shown that in regions with threshold fine structure the ability to detect amplitude modulations on a tonal carrier strongly depends on the position of the carrier relative to the fine structure, even for low modulation frequencies: MDTs are higher when the carrier falls at a fine-structure minimum than when it falls at a fine-structure maximum (Zwicker (1986); Long (1993); Heise *et al.* (2009)). In other words, frequency regions that are more sensitive to tones are less sensitive to AM. This effect of carrier position on MDTs exists regardless of whether the comparison is made at equal carrier sound pressure levels (SPL) or sensation levels (SL) (Heise *et al.* (2009)). That is, in regions with fine structure, the detectability of an AM is not directly determined by the loudness of the carrier, since at an equal carrier SPL the stimuli at fine-structure minima are perceived as louder than those at fine-structure maxima (Mauermann *et al.*, 2004).

This study aims at identifying possible mechanisms behind the different modulation detection performances at fine-structure minima and maxima. In literature three explanations have been offered, which are outlined below (Long, 1993; Heise *et al.*, 2009). Two of these attribute the observed effect to interactions of the stimulus with otoacoustic emissions (OAEs). A close link between spontaneous otoacoustic emissions (SOAEs) or large evoked otoacoustic emissions (EOAEs) and threshold fine structure has been established both experimentally (Zwicker and Schloth (1984); Long and Tubis (1988); McFadden and Mishra (1993)) and in cochlear models (Zweig and Shera (1995); Talmadge *et al.* (1998)) (for an overview, see Johnson *et al.* (2008)). SOAEs or large EOAEs[1] tend to occur at fine-structure minima. From a model point of view both phenomena originate from the same cochlear mechanisms: Different cochlear models incorporate a mechanism that leads to a coherent reflection of any incoming sound at randomly distributed mechanical inhomogeneities in the region of maximum cochlear excitation (Zweig and Shera, 1995; Talmadge *et al.*, 1998). According to these models, the reflected sound travels to the stapes where, due to the impedance mismatch, some of it is reflected back into the cochlea again (e.g., Shera and Zweig, 1993). The returning reflection will either enhance or partially cancel energy at the original cochlear reflection site,

[a)]Portions of this work were presented at the 32nd Midwinter Meeting of the Association for Research in Otorhinolaryngology, Baltimore, MD, February 2009 and at the 35th Meeting of the German Acoustical Society, Rotterdam, Netherlands, March 2009.

[b)]Author to whom correspondence should be addressed. Electronic mail: stephan.heise@uni-oldenburg.de

depending on the round-trip travel time. If the sound is not canceled by the returning echo and if it was reflected the first time, it will be reflected again unless the properties of the cochlea have changed. Multiple internal reflections of cochlear traveling waves will occur, generating a resonance pattern in the cochlea (Zweig and Shera, 1995). These resonances enhance the response of the basilar membrane to sounds at some frequencies, and reduce its response to sounds at others, thus resulting in threshold fine-structure minima and maxima, respectively. The same resonance behavior can be used to explain in general the pseudoperiodicity observed in all types of OAEs that originate from a single source, such as SOAEs. Therefore cochlear resonances (CRs) provide a common explanation for threshold fine structure and these types of OAEs (see Talmadge *et al.*, 1998).

Apparently not all sound that is generated in the cochlea can be measured in the ear canal as a SOAE or a large EOAE. This is indicated by the fact that not every fine-structure minimum is associated with a SOAE. In particular, strong CRs may result from a high reflectivity of the oval window, which in turn keeps most of the sound inside the cochlea. Therefore the term CR will be used instead of the term "SOAE" throughout this article to refer to both SOAEs as well as other similar resonances, which may be reflected in large EOAEs and/or minima in the threshold in quiet.

A first explanation for the different AM sensitivity at fine-structure minima and maxima is based on an effect known as "monaural diplacusis." Many people report that tones near the threshold in quiet sometimes sound rough, which can be identified as the beating of a CR with the external stimulus (Long and Tubis, 1988). Such a beating could to some extent mask the modulation that is supposed to be detected. If the frequency of a fine-structure minimum is not measured exactly, then the frequency difference between the associated CR and the carrier of the AM signal may be larger than the region of entrainment for the CR, and beating may occur (Long and Tubis, 1988). This would lead to higher MDTs at fine-structure minima than at maxima. This hypothesis will be denoted as $\mathcal{H}_{\text{beat}}$ in the following.

A second explanation assumes that CRs increase MDTs at fine-structure minima in the following way: When the carrier of an AM signal falls at a fine-structure minimum it is likely to entrain the CR, so that the amplitudes of the synchronized CR and the stimulus add in phase. This would reduce the effective modulation depth at fine-structure minima and make the modulation harder to detect (Long, 1993). In the following this hypothesis will be denoted as $\mathcal{H}_{\text{sync}}$.

A third explanation is based on the assumption that threshold fine structure arises from variations in the gain of the cochlear amplifier with frequency. The gain is higher at fine-structure minima than at maxima. Such a frequency-specific gain function can be implemented as a filter whose shape is the inverse of the fine structure. So when the carrier of an AM signal is at a fine-structure minimum, the carrier receives a higher gain than the sidebands. Effectively this would reduce the modulation depth of the stimulus and lead to relatively high MDTs. Likewise the effective modulation depth would be increased for stimuli at fine-structure maxima, which would result in relatively low MDTs. This hypothesis will be denoted as $\mathcal{H}_{\text{factor}}$ in the following and assumes a level-independent gain, i.e., a linear filter. While this mechanism cannot account for the whole effect (Heise *et al.*, 2009), it may still add to one of the aforementioned mechanisms.

In a nutshell, the three hypotheses assume a modification of the stimulus in the cochlea either by the addition of a CR at a different frequency ($\mathcal{H}_{\text{beat}}$) or by the addition of a CR in phase ($\mathcal{H}_{\text{sync}}$) or the multiplication with a gain factor ($\mathcal{H}_{\text{factor}}$). To probe these basic concepts three experiments are carried out in this study. All vary a different stimulus parameter while comparing MDTs at fine-structure minima and maxima. The experiments are also carried out in regions without threshold fine structure. This reference allows one to characterize fine-structure specific changes in the perception of AM. Further, MDTs at fine-structure minima are simulated on the basis of $\mathcal{H}_{\text{sync}}$ in order to probe this hypothesis quantitatively. In addition, SOAEs are measured.

In experiment 1, temporal modulation transfer functions (TMTFs) in which the modulation frequency is varied are measured for tonal carriers in order to test whether the position of the sidebands relative to the fine structure has an influence on the MDTs. The modulation frequencies are fitted individually to span one fine-structure cycle. An influence of the modulation frequency is expected on the basis of $\mathcal{H}_{\text{factor}}$ since the relation between the levels of the carrier and the sidebands after filtering critically depends on the position of the sidebands, i.e., on the modulation frequency. In contrast, $\mathcal{H}_{\text{sync}}$ predicts a change in effective modulation depth, which only depends on the carrier frequency so that the MDTs should be independent of modulation frequency. The modulation masking proposed by $\mathcal{H}_{\text{beat}}$ would lead to a TMTF that is similar to a modulation masking pattern where the beating masks the target modulation.

The dependence of the MDTs on stimulus level is examined in experiment 2 by measuring MDT growth functions for tonal carriers. From $\mathcal{H}_{\text{sync}}$ the growth functions at fine-structure minima would be expected to be steeper than those at fine-structure maxima, because the added component in minima would become less effective toward higher levels. At high levels MDTs at fine-structure minima and maxima should be the same. Similar growth functions are predicted by $\mathcal{H}_{\text{beat}}$ since also the beating becomes weaker toward higher levels. On the other hand, a difference in MDTs based on different gain factors ($\mathcal{H}_{\text{factor}}$) should be unaffected by stimulus level.

The level dependence of the MDTs is also analyzed using a narrow-band noise carrier (experiment 3). The inherent envelope fluctuations of the non-deterministic carrier should mask the beating proposed in $\mathcal{H}_{\text{beat}}$. That is, for this type of carrier, an additional beating should hardly affect modulation detection performance. Thus MDTs at fine-structure minima and maxima should be very similar for a narrow-band noise carrier. For $\mathcal{H}_{\text{factor}}$ a difference in MDTs at minima and maxima should remain for the narrow-band noise carrier, since the difference between the gain for the carrier and the sidebands still exists. However, the effect is expected to be

reduced due to the broader spectral components. The effect should also persist under $\mathcal{H}_{\text{sync}}$ although it may be reduced if the synchronized CR is not able to follow the instantaneous frequency of the noise carrier, so that instead of the amplitudes the energies of the carrier and the CR add.

II. METHODS

A. Subjects

15 subjects (4 females and 11 males) with ages ranging from 18 to 35 years participated in the experiments. All had absolute thresholds lower than 15 dB hearing loss (HL) at all audiometric frequencies, except for subjects HS and OM who had a threshold of 20 dB HL at 6 kHz.

B. Setup

The subjects were seated in a double-walled, sound-attenuating booth. For the psychoacoustic measurements the signals were generated digitally on a PC in MATLAB at a sampling rate of 44.1 kHz and fed to an RME ADI8Pro digital-to-analog converter via an RME Digi96/8 PAD sound card. The converted signal was amplified by a Tucker Davis HB7 amplifier and played back via Sennheiser HDA200 headphones. The headphones were coupler-calibrated on a Brüel & Kjær artificial ear (type 4153). SOAEs were recorded digitally at a sampling rate of 48 kHz with an Etymotic Research ER10C insert microphone in the closed ear canal, which was connected to the same RME ADI8Pro analog-to-digital converter as in the psychoacoustic setup.

C. Threshold in quiet measurements

Thresholds in quiet were measured in two ways: with a tracking procedure (FINESS) for measuring threshold fine structure with a high frequency resolution, and with an adaptive three-alternative forced-choice procedure (3-AFC) for obtaining an unbiased threshold estimate at a limited number of frequencies.

For a detailed description of the FINESS procedure the reader is referred to Heise *et al.*, 2008. FINESS was developed for reliably screening threshold fine structure while at the same time keeping the acquisition time as low as possible. The shape of threshold fine structure is measured very accurately by the FINESS method. Since it is a tracking procedure the absolute threshold values may be biased by the subject's internal threshold criterion (Heise *et al.*, 2008). In order to correct for this potential bias the thresholds obtained by the FINESS method are shifted vertically so as to minimize their squared distance to the thresholds obtained by the 3-AFC procedure.

In the adaptive 3-AFC procedure the subjects were presented with three 250-ms intervals, one of which contained the stimulus. The intervals were graphically indicated, and separated by 500 ms of silence. Subjects had to indicate which interval contained the stimulus. Visual feedback was provided after each trial. A one-up, two-down stepping rule was applied, which estimates the level required for 70.7% correct responses (Levitt, 1971). The stimulus level started at 15 dB SPL and was changed in 6-dB steps. The step size was reduced to 3 dB after two reversals and to 1 dB after four reversals. Eight reversals were measured at the smallest step size and their levels averaged to give a threshold estimate.

D. Modulation detection threshold measurements

MDTs of a sinusoidally amplitude modulated tone were obtained by means of an adaptive 3-AFC procedure with a one-up, two-down stepping rule. Subjects were presented with three tones, one of which was amplitude modulated. The stimuli had a duration of 500 ms including 20 ms raised-cosine rise/fall ramps, and were separated by 300 ms of silence. The SPL of the carrier component was equal across the three intervals. As a tracking variable the modulation depth in decibel ($20 \log_{10} m$) was used. It started at -4 dB and was changed in steps of 4 dB. The step size was reduced to 2 dB after two reversals and to 1 dB after four reversals. Another eight reversals were measured and averaged to give a threshold estimate. The modulation depth was restricted to values less than or equal to 0 dB. If a subject could not detect the modulated stimulus at this maximum modulation depth three times within the same run, the run was aborted. In order to assess the relevance of beating between the stimulus and a CR the subjects were asked to report a possible beating after each trial: When more than one stimulus in the trial sounded modulated they were instructed to give their response on an alternative set of keys ({7,8,9} instead of {1,2,3}).

E. SOAE measurements

SOAEs were recorded digitally using the setup described above. Further online analysis was performed using custom-made measurement software based on MATLAB. The online analysis included artifact rejection, Fourier transformation, averaging in the frequency domain (in order to reduce the variability of the noise floor), and automatic SOAE detection. Peaks in the averaged power spectrum were detected as a SOAE if they exceeded the local noise floor in a region of ± 10 Hz around the peak (at a frequency resolution of 0.5 Hz) by more than two standard deviations.

F. Procedure

In a first step the threshold in quiet was screened for possible fine structure over one to two octaves in the range between 1 and 4 kHz. For this purpose the FINESS procedure was used with the default frequency resolution of 100 frequencies per octave.[2] Then a region of about 1/3 octave was selected for the modulation detection experiments. In ten subjects this was a region with threshold fine structure, and in five subjects it was a region without fine structure. In this region the threshold in quiet was re-measured at least three times using the FINESS method with a higher frequency resolution of 150 frequencies per octave. The individual threshold curves and the averaged curve for each subject are shown in Fig. 1 as lines (thin gray and thick black, respectively). A measure for the local amount of fine structure in the mean threshold curve was obtained by the FINESS-detector algorithm (Heise *et al.*, 2008).[3] This measure, coded in gray scales, is shown in a bar at the top of each panel in

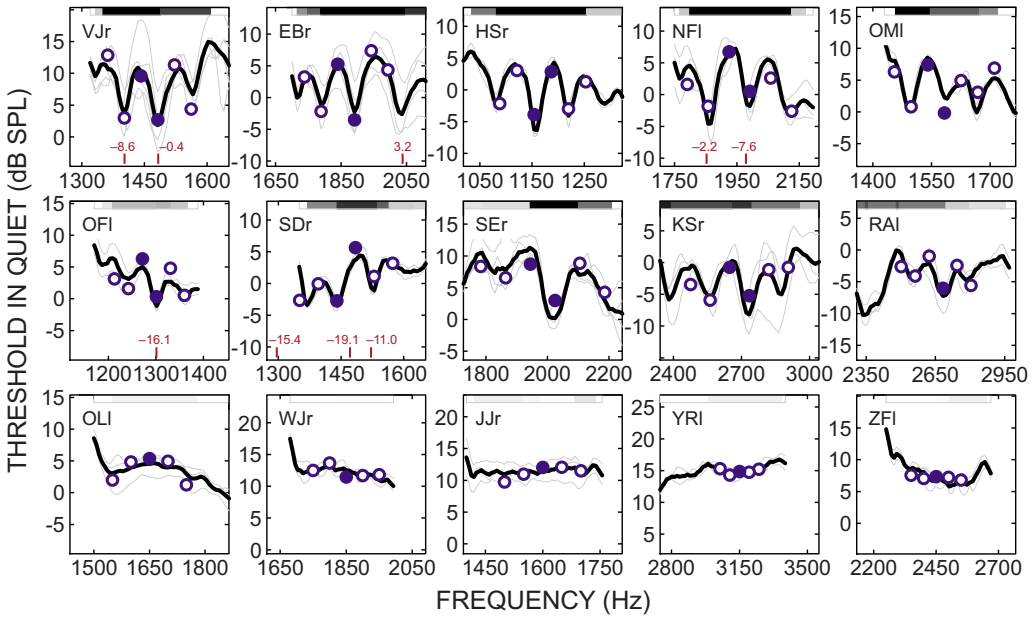


FIG. 1. (Color online) Individual thresholds in quiet. Thresholds from the FINESS procedure are displayed as lines (thin: individual runs, thick: mean data). Thresholds from the 3-AFC procedure are displayed as open (○) or filled (●) circles (mean data across all sessions). Filled circles indicate frequencies that were used as carrier frequencies in the modulation detection experiments. For subjects that showed SOAEs, their frequencies are indicated as small vertical lines at the bottom of each panel; the SOAE level in dB SPL is given as a number above the line. A bar at the top of each panel shows the local amount of fine structure coded as gray scales (black: very pronounced fine structure, white: no fine structure).

Fig. 1 where black indicates very pronounced fine structure (more than 6 dB_{avg}[3]) and white indicates regions without fine structure (0 dB_{avg}). From the mean threshold curve a fine-structure minimum and an adjacent maximum were chosen whose frequencies f_{car}^{min} and f_{car}^{max} served as carrier frequencies in the modulation detection experiments. Since fine-structure minima tend to be sharper than maxima, the minimum was selected first and the frequency of the maximum was then defined as $f_{car}^{max} := \Delta f/2$, where Δf is the frequency difference between the minimum at f_{car}^{min} and an adjacent minimum. For subjects without fine structure, only one frequency f_{car}^{flat} was chosen, and Δf was set to 100 Hz, which is the average Δf of the subjects with fine structure. The carrier frequencies are marked as filled circles in Fig. 1. SOAEs were measured for all subjects. SOAEs that were more than 3 dB above the noise floor are shown in Fig. 1 if they fell into the displayed frequency range. Their frequencies are indicated by small vertical lines at the bottom of each panel and their SPLs are given next to the lines.

After these initial preparations three to seven sessions followed in which the actual modulation detection experiments were carried out. In each of these sessions the pure-tone thresholds in quiet at the carrier frequencies and at adjacent fine-structure extreme values that were of interest for the specific experiment were measured with an adaptive 3-AFC procedure prior to measuring MDTs. The threshold in quiet was measured three times at the carrier frequencies and twice at the other frequencies. The mean threshold data from all sessions are plotted as circles in Fig. 1. Note that in this and following figures the subjects are sorted according to their local amount of fine structure, which was calculated as the mean absolute difference between thresholds in quiet at adjacent fine-structure extreme values as measured with the 3-AFC procedure. MDTs were measured with a 3-AFC procedure as described in Sec. II D. 4 threshold estimates were obtained for each experimental condition and in most cases 16–20 estimates were acquired per session.

In experiment 1, TMTFs were measured for tonal carriers at fine-structure minima and maxima and in regions without fine structure. The modulation frequencies—i.e., the positions of the spectral sidebands—were adapted to the subject's individual fine structure in the threshold in quiet. Expressed in fractions of the frequency difference Δf between two adjacent fine-structure minima, the modulation frequencies were f_{mod}=8 Hz, $\frac{1}{4}\Delta f$, $\frac{1}{2}\Delta f$, $\frac{3}{4}\Delta f$ and Δf. At the lowest modulation frequency the sidebands were close to the carrier. 8 Hz was chosen as a minimum modulation frequency in order to provide the subjects with at least four modulation cycles for detecting the modulation [as has been suggested by Lee and Bacon (1997)]. For $f_{mod}=\Delta f/2$ the sidebands fell close to fine-structure maxima when the carrier was at a fine-structure minimum and close to minima when the carrier was at a maximum. For $f_{mod}=\Delta f$ the sidebands fell close to fine-structure extreme values of the same type as the one at the carrier. The carrier could either fall at a fine-structure maximum (f_{car}^{max}) and have a level of 15 dB SL (condition max), or it could fall at a fine-structure minimum (f_{car}^{min}) where it could either have the same absolute level as in the max condition (condition minSPL) or a level of 15 dB SL (condition minSL). In subjects without fine structure only one carrier at f_{car}^{flat} was used with a level of 15 dB SL (condition flat).

In experiment 2, MDTs were measured as a function of level for tonal carriers at fine-structure minima and maxima and in regions without fine structure. MDTs were obtained for five carrier levels (L_{car}=7.5, 15, 22.5, 30, and 37.5 dB SL) and one modulation frequency ($f_{mod}=\Delta f/2$). As in ex-

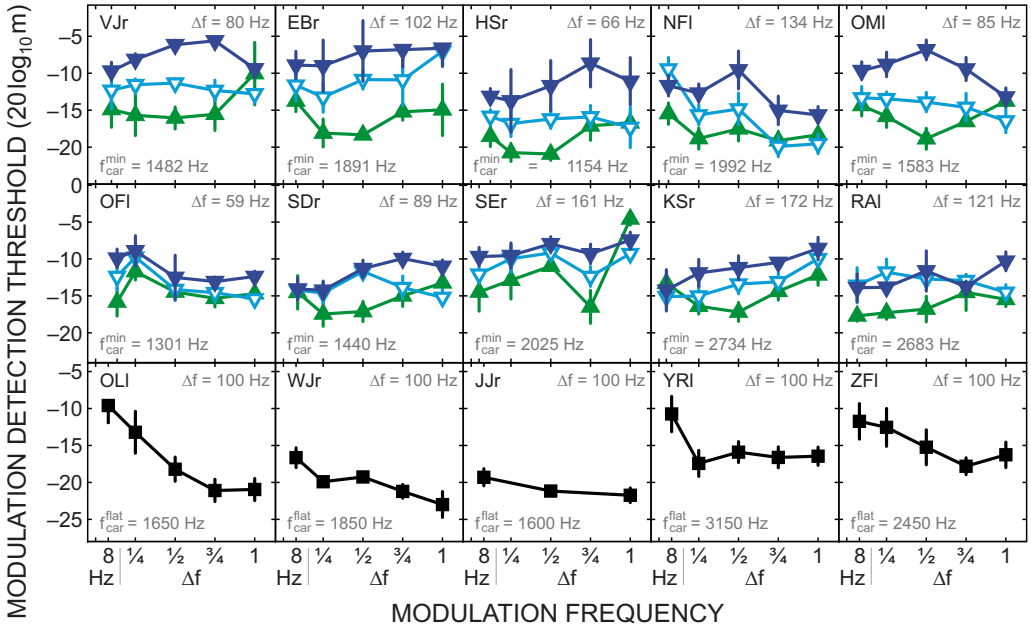


FIG. 2. (Color online) Individual results from experiment 1. TMTFs are shown for the conditions max (▲), minSPL (▽), minSL (▼), and flat (■). Every data point is the mean of four threshold estimates; error bars give standard deviations. The values for Δf—i.e., the highest modulation frequency—and for one carrier frequency are given.

periment 1, the three positions of the carrier relative to the individual fine structure are referred to as conditions max, minSL, and flat.

In experiment 3, MDTs were measured for narrow-band noise carriers at fine-structure minima and maxima and in regions without fine structure. The carriers had a bandwidth of 32 Hz and were generated by setting the frequency components of a Gaussian noise signal to zero outside the desired passband. A bandwidth of 32 Hz was chosen in order to ensure a sufficient masking of potential beating near fine-structure minima while at the same time limiting any masking of the test modulation as well as any reduction in fine structure produced by the broadening of the signal (cf. Long and Tubis, 1988). MDTs were obtained for two carrier levels (L_{car}=15 and 30 dB SL) and one modulation frequency ($f_{mod}=\Delta f/2$). The three positions of the carrier relative to the individual fine structure are referred to as conditions max, minSL, and flat as in experiments 1 and 2. Unlike in experiments 1 and 2, subjects were not asked to report any beating in this experiment since potential beating would have been very hard to discriminate from the inherent fluctuations of the noise carrier.

III. RESULTS

A. Experiment 1: TMTFs

The individual results are shown in Fig. 2. As in Fig. 1 the subjects are sorted according to their local amount of threshold fine structure, i.e., the first row shows subjects with strong fine structure, the second row shows subjects with moderate fine structure, and the third row shows subjects without fine structure. The panels contain the mean results from the TMTF measurements in the conditions max (upward-pointing triangles), minSPL (downward-pointing open triangles), minSL (downward-pointing filled triangles), and flat (squares). The values of Δf and of f_{car}^{min} or f_{car}^{flat} are given for each subject.

In general, the TMTFs from fine-structure regions show the following consistent differences between experimental conditions: TMTFs are lowest in the max condition, highest in the minSL condition, and in-between in the minSPL condition. The effect of threshold fine structure on modulation detection performance may be quantified by the difference between the MDTs in the max and the minSL conditions. This effect is largest around $f_{mod}=\Delta f/2$, i.e., when the sidebands were close to fine-structure extreme values adjacent to the carrier frequency. The detailed results vary considerably between subjects, which suggests a strong influence of the individual fine structure on the MDTs. This view is supported by the difference between the MDTs in the max and the minSL condition at $f_{mod}=\Delta f/2$ being correlated with the local amount of fine structure in the ten subjects showing fine structure (correlation coefficient=0.72). Here, the local amount of fine structure was calculated as the mean absolute difference of thresholds in quiet at fine-structure extreme values as measured in the TMTF sessions.[4]

To average the data the subjects were grouped into three groups according to their average amount of fine structure in all sessions (i.e., groups correspond to rows in Fig. 2). The mean results of the three groups are shown in the left column in Fig. 3. The differences in shape and absolute value between the TMTFs at fine-structure minima and maxima are clearly larger for the group of subjects with strong fine structure (top panel) than for the group of subjects with moderate fine structure (middle panel). In the group with strong fine structure the TMTF in the minSPL condition is nearly constant, whereas the modulation frequency seems to have opposite effects on the MDTs in conditions max and minSL: In

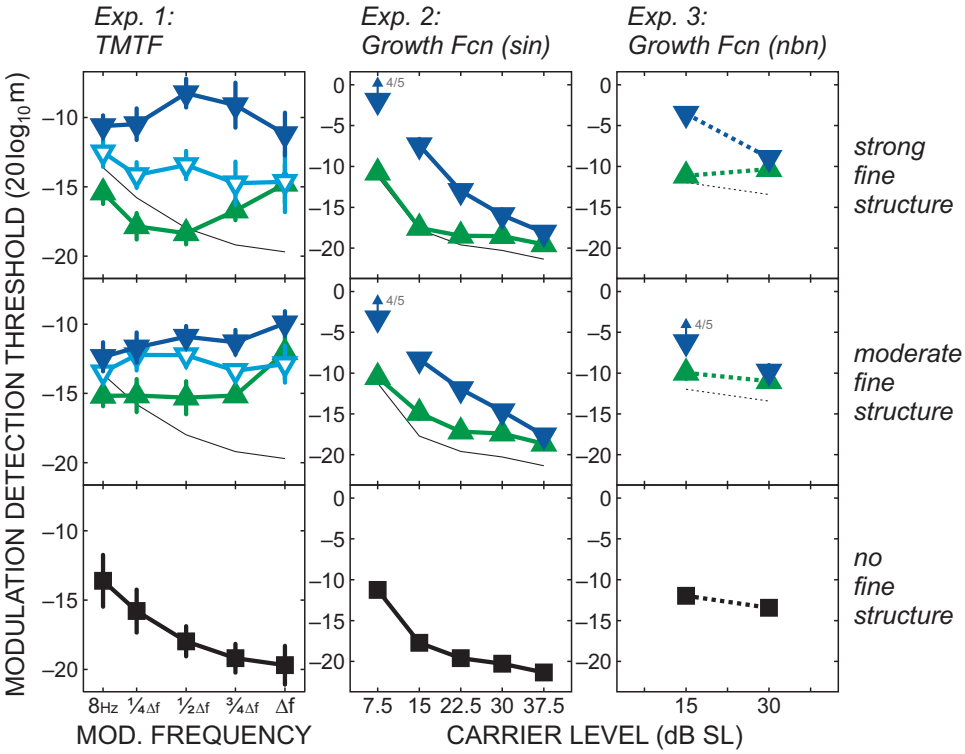


FIG. 3. (Color online) Mean results for all experiments. Each column represents results from one experiment. For averaging purposes the subjects were grouped according to their local amount of fine structure: strong (first row), moderate (second row), or no fine structure (third row). Each data point is the mean across five subjects, except for the cases that are indicated with upward-pointing arrows. The numbers next to the arrow give the proportion of subjects of which at least one threshold estimate could be obtained and which were included in the mean. Error bars give standard errors (often hidden behind the symbol). Different symbols distinguish between experimental conditions: max (▲), minSPL (▽), minSL (▼), and flat (■). The data from the bottom panels are redrawn in the upper panels as thin lines without symbols in order to facilitate the comparison between fine-structure and reference data.

the max condition MDTs are relatively high for the lowest modulation frequency, improve toward $f_{\mathrm{mod}}=\Delta f/2$, and increase again toward $f_{\mathrm{mod}}=\Delta f$, whereas in the minSL condition the TMTF peaks at $f_{\mathrm{mod}}=\Delta f/2$. The mean results from the group with moderate fine structure still show an effect of fine structure on the MDTs but this hardly depends on modulation frequency. The reference data from the flat condition (bottom panel) is redrawn in the upper two panels as a thin black line. The comparison of the results from all conditions shows that the TMTFs from regions with fine structure mostly lie above the average TMTF from regions without fine structure. This indicates that the main effect of threshold fine structure in this context can be characterized as a reduction in sensitivity to amplitude modulation. This reduction is large when the carrier is at a fine-structure minimum while the MDTs in the max condition only show clear deviations from the MDTs in the flat condition at high modulation frequencies.

B. Experiment 2: MDT growth functions with tonal carriers

The individual results for the conditions max (upward-pointing triangles), minSL (downward-pointing triangles), and flat (squares) are shown in Fig. 4. The modulation frequency was $\Delta f/2$ in all conditions. The circles show simulations that are described in Sec. IV. In some runs at low carrier levels subjects were not able to detect the modulation even at $m=100\%$. Averaged data points including such runs are not connected to other data points and are highlighted by upward-pointing arrows, indicating that the MDT may actually be higher than shown. The ratio of valid to total runs is given next to the arrow.

The MDTs in all conditions decrease as level increases. The difference between MDTs at fine-structure minima and maxima is largest at the lowest carrier level and gradually disappears toward higher carrier levels. The interindividual differences in the results are much smaller than in the TMTF experiment. For the ten subjects with strong and moderate fine structures the difference of MDTs in the max and the minSL condition at $L_{\mathrm{car}}=15$ dB SL may be defined as a measure of the effect of fine structure on MDTs—15 dB SL is the lowest carrier level at which complete data of all subjects exist. This effect is highly correlated (correlation coefficient $\rho=0.85$) with the local amount of fine structure, which was quantified by the mean absolute difference of thresholds in quiet at adjacent fine-structure extreme values.

The mean data averaged across subjects with strong, moderate, and no fine structure are shown in the middle column of Fig. 3. Here the numbers next to upward-pointing arrows indicate how many subjects had at least one valid measurement that could be included in the average. The MDTs in the max condition deviate only slightly from the MDTs in the flat condition, while—as in the TMTFs—the deviation is much stronger for the MDTs in the minSL condition.

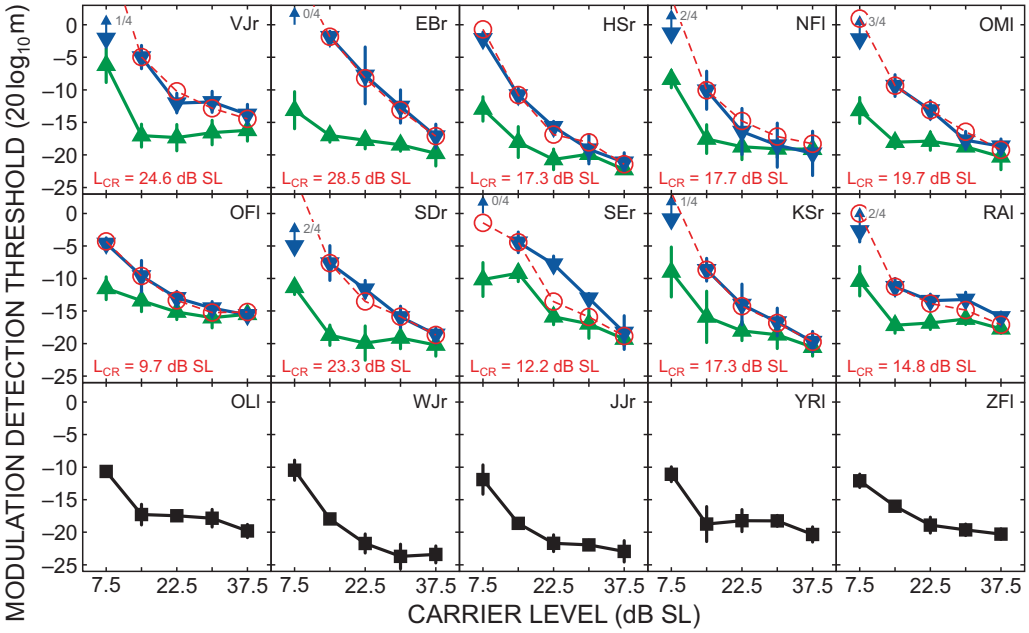


FIG. 4. (Color online) Individual results for MDT growth functions with tonal carriers in the conditions max (▲), minSL (▼), and flat (■). Data points without arrows represent the mean of four threshold estimates; error bars give standard deviations. Arrows indicate that not all four threshold estimates could be obtained; the numbers next to the arrows give the ratio of valid to total runs. Circles (○) show simulations of the data in the minSL condition on the basis of the hypothesis $\mathcal{H}_{\text{sync}}$ (see Sec. IV for details). The levels of the CRs that were used for the simulation are given in the lower left corner of each panel.

C. Experiment 3: MDT growth functions with narrow-band noise carriers

The individual results for the conditions max (upward-pointing triangles), minSL (downward-pointing triangles), and flat (squares) are shown in Fig. 5. As before, arrows connected to data points indicate that in some runs the subject could not detect the modulation even at $m=100\%$. Similar to the results for tonal carriers, the MDTs in the minSL condition are higher than in the max condition at the lower carrier level. At the higher carrier level the data of the two conditions coincide, which is mainly due to the MDTs in the minSL condition decreasing with level. The MDTs in the max condition show no clear trend across levels. The data in the flat condition tend to decrease slightly with level.

The mean data are shown in the right column of Fig. 3. Compared to the corresponding results for tonal carriers, the results for narrow-band noise carriers are shifted by about 6 dB toward higher MDTs. In addition, the difference between MDTs in the max and the minSL conditions at 15 dB SL is about 3 dB smaller for the narrow-band noise carriers than for tonal carriers. The relation between MDTs from regions with and without fine structure is consistent with that in experiment 2.

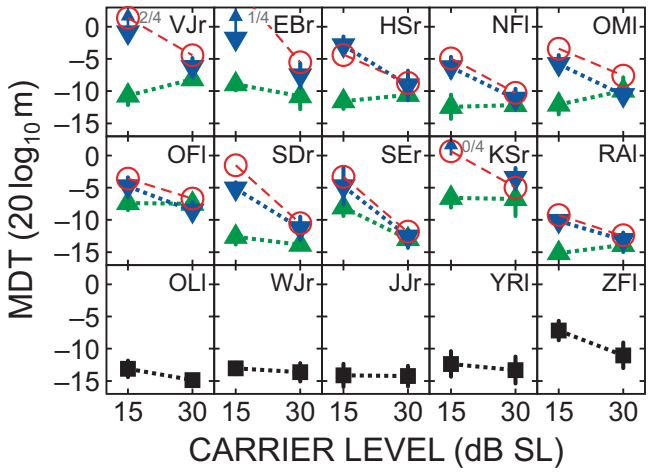


FIG. 5. (Color online) Individual results for MDT growth functions with narrow-band noise carriers in the conditions max (▲), minSL (▼), and flat (■). Data points without arrows represent the mean of four threshold estimates; error bars give standard deviations. Arrows indicate that not all four threshold estimates could be obtained; the numbers next to the arrows give the ratio of valid to total runs. Circles (○) show simulations of the data in the minSL condition on the basis of the hypothesis $\mathcal{H}_{\text{sync}}$ (see Sec. IV for details).

IV. DISCUSSION

MDTs have been measured as a function of modulation frequency, carrier level, and carrier bandwidth. The results are summarized in Fig. 3. In the following the experimental results are discussed with respect to various aspects including other experimental data and, in particular, the three possible mechanisms underlying the modification of MDTs by threshold fine structure that were introduced in Sec. I.

A. Comparison with literature data

So far, three studies have dealt with modulation perception in connection with threshold fine structure. Zwicker (1986) used a tracking procedure to measure MDTs at several fine-structure minima and maxima of four subjects for two carrier levels ($L_{\text{car}}=7$ and 17 dB SL) and four different modulation frequencies ($f_{\text{mod}}=1,4,16,64$ Hz). Long (1993) compared MDTs at fine-structure minima and maxima for equal carrier SPL ($L_{\text{car}}=15,20,\ldots,60$ dB SPL) at a modulation frequency of 4 Hz in three subjects. Heise *et al.* (2009) measured MDTs in the same three conditions max, minSPL

and minSL as in experiment 1 (L_{car}=15 dB SL in the conditions max and minSL, and in the minSPL condition the SPL was the same as in the max condition) for $f_{mod}=\Delta f/2$ in nine subjects.

Qualitatively, all studies agree on MDTs at fine-structure minima being higher than MDTs at fine-structure maxima. However, some quantitative differences between the results exist, which are probably due to differences in stimulus parameters, experimental methods, and subjects. The absolute MDTs in the max and the minSL conditions in the present study are 2–4 dB higher than the data in Zwicker (1986), but agree well with the data in Long (1993) and in Heise *et al.* (2009). The differences in MDTs at fine-structure minima and maxima in the present study are consistent with mean data from Zwicker (1986) and from Heise *et al.* (2009). In contrast, Long (1993) observed about 3 dB smaller effects for levels below 25 dB SL and about 3 dB larger effects for levels above 30 dB SL.

Concerning the influence of modulation frequency and carrier level, Zwicker (1986) found no dependence of MDTs on modulation frequency, which agrees with the results from experiment 1 for the subjects with moderate fine structure. The MDT growth functions from experiment 2 are basically in line with the data in Zwicker (1986) and Long (1993), although the growth functions from fine-structure minima and maxima merge at somewhat higher levels in Long (1993) than in the present study.

TMTFs in the flat condition show an initial decrease at low modulation frequencies before becoming constant for higher modulation frequencies. Literature TMTFs are rather heterogeneous in the according modulation-frequency range ($f_{mod} \leq 100$ Hz): Some studies find a steadily increasing TMTF (Kohlrausch *et al.* (2000), data for f_{car}=1 kHz and $L_{car} \leq 30$ dB SPL; Millman and Bacon (2008), data for L_{car}=10 dB SL), others report a flat TMTF (Moore and Glasberg (2001), data for L_{car}=30 dB SPL) which is similar to the TMTF at high levels (Strickland and Viemeister, 1997; Kohlrausch *et al.*, 2000). The decreasing TMTF found in the present study is very similar to the TMTF in Millman and Bacon, 2008 (data for L_{car}=30 dB SL), but is also observable in data from Yost and Sheft, 1997 at high levels, at least up to f_{mod}=16 Hz. An initial decrease may be related to the number of modulation cycles available to the subject for detecting the modulation, even though a minimal number of four cycles (equivalent to f_{mod}=8 Hz at a duration of 500 ms) has been reported for a stable AM discrimination performance (Lee and Bacon, 1997). The MDT growth functions in the flat condition agree with mean data from Kohlrausch *et al.*, 2000 for f_{mod}=50 Hz and with mean data in Millman and Bacon, 2008 for f_{mod}=64 Hz and L_{car}=30 dB SL. The data in Millman and Bacon, 2008 for L_{car}=10 dB SL lie between the data for the flat and the minSL conditions in the present study.

Little is known about MDTs for noise carriers with bandwidths comparable to that used here and at levels near the threshold in quiet. Dau *et al.* (1997) measured MDTs for a 31-Hz wide Gaussian noise centered at 5 kHz for a level of 65 dB SPL. At f_{mod}=50 Hz they found MDTs around −14 dB, which, despite the level difference, agrees well with the MDTs in the flat condition in the present study. This indicates that already for carrier bandwidths around 30 Hz, MDTs depend far less on carrier level than MDTs for tonal carriers. The 6-dB difference between MDTs for tonal and narrow-band noise carriers in the present study is somewhat smaller than in studies by Dau *et al.* (1997, 1999). This can be ascribed to the MDTs for tonal carriers being lower in literature because of the higher level (e.g., Kohlrausch *et al.*, 2000).

B. The role of spectral cues

Literature on TMTFs at high levels suggests that modulation detection is based on spectral cues for high modulation frequencies that exceed the bandwidth of the auditory filter at the carrier frequency (Strickland and Viemeister, 1997; Kohlrausch *et al.*, 2000; Moore and Glasberg, 2001). That is, listeners can discriminate the modulated from the unmodulated stimulus by listening for the presence of the sidebands rather than detecting the temporal fluctuations in the envelope (Yost and Sheft, 1997). Since there is evidence that the auditory filter width decreases toward low levels (Glasberg and Moore, 2000)—and particularly at CRs (Long, 1984; Bright, 1997)—one may expect to find effects of spectral cues at lower modulation frequencies at low levels than at high levels. However, even if the sidebands fall at frequencies where they could in principle be resolved, at low levels they most often fall below the threshold in quiet (Kohlrausch *et al.*, 2000). In the present data only about 15% of the sidebands at MDT are above the corresponding threshold in quiet, mostly in the minSPL condition. The mean sideband level at MDT is −4.7 dB SL. These findings make it unlikely that spectral cues play an important role in the present data.

C. Comparing MDTs from regions with and without fine structure

MDTs from regions with threshold fine structure were found to generally be higher than MDTs from regions without fine structure. This effect is largest at fine-structure minima whereas the MDTs at fine-structure maxima are often very similar to the MDTs from regions without fine structure. This observation is supported by the following analysis of the MDT growth functions from experiment 2. In general, MDTs for band-limited carriers improve with increasing level, which is sometimes called the "near miss to Weber's law" (e.g., Wojtczak and Viemeister, 2008). Quantitatively the near miss can be described by the slope of the data when plotted in terms of $(20 \log_{10}(2A\sqrt{m}))$ versus $(20 \log_{10}(A(1-m)))$, where A is the amplitude of the unmodulated signal and m is the modulation depth.[5] In band-limited carriers this slope is found to be near 0.9 instead of 1 as would be expected from Weber's law. The present data show slopes of 0.92 (max condition), 0.67 (minSL condition), and 0.91 (flat condition). These values are mean slopes from linear fits to the individual data in which the data for L_{car}=7.5 dB SL were disregarded because most subjects had some invalid runs at this carrier level. That is, the data for the max and the

TABLE I. Percentage of AFC trials for which the subjects reported to have heard a beating between the stimulus and a CR in the modulation detection experiments. Numbers are given for carriers at fine-structure minima and maxima separately. No beating was reported by the five subjects for which the carrier fell in a region without fine structure.

Subject	VJr	EBr	HSr	NFl	OMl	OFl	SDr	SEr	KSr	RAl
Beats at f_{car}^{min}	28	15	2	25	6	0	0	59	10	9
Beats at f_{car}^{max}	32	4	9	9	4	0	1	51	7	7

flat conditions show the near miss to Weber's law whereas the data for the minSL condition deviate much more strongly from Weber's law.

D. The concept of beating with CRs ($\mathcal{H}_{beat}$)

As a first possible mechanism behind the higher MDTs at fine-structure minima compared to those at fine-structure maxima a masking of the test modulation by the beating of a CR with the stimulus carrier at fine-structure minima is examined. When using a narrow-band noise carrier the masking of the test modulation by the inherent amplitude modulations of the noise carrier should dominate over the masking by the beating. This should lead to very similar MDTs at fine-structure minima and maxima. However, the MDTs in experiment 3 still show a clear difference between conditions at fine-structure minima and maxima. Thus $\mathcal{H}_{beat}$ may be rejected as the only underlying mechanism for the difference of MDTs at fine-structure minima and maxima. In agreement, the frequencies of the SOAEs that were measured (see Fig. 1) do not differ from the frequencies of associated fine-structure minima by more than 7 Hz, which makes it seem unlikely that the carriers at f_{car}^{min} were outside the entrainment region of the associated CRs (cf. Long and Tubis, 1988).

Nevertheless beating may play a role in modulation detection near threshold. Eight of the ten subjects with fine structure reported trials in which the modulation detection was disturbed by a beating in at least one of the reference intervals (see Table I). There are large variations between subjects regarding the number of such trials and at which modulation frequencies they tended to occur. Interestingly, almost as many beat trials were reported at fine-structure maxima as were at fine-structure minima. Hence the beating of a CR with the stimulus carrier may be a disturbing factor for modulation detection near threshold in general. This is in line with the finding that most MDTs in the max condition are higher than the corresponding MDTs in the flat condition in the present study.

E. The concept of entrained CRs ($\mathcal{H}_{sync}$)

As another possible mechanism, the reduction in effective modulation depth by the in-phase addition of a CR to the stimulus carrier at fine-structure minima was proposed. Such a reduction should be independent of modulation frequency and decrease with increasing carrier level. Consistently, the TMTFs at fine-structure minima are higher than at fine-structure maxima and higher than in regions without fine structure, and the TMTFs in the minSPL condition are lower than in the minSL condition. The TMTFs in subjects with moderate fine structure are nearly flat, which is in agreement with $\mathcal{H}_{sync}$. However, subjects with strong fine structure show a dependency of MDTs on modulation frequency in the conditions max and minSL. In the minSPL condition—where the carrier level is 3–10 dB higher than in the minSL condition, so that the fine structure becomes less effective (Mauermann *et al.*, 2004)—the TMTF is almost flat as in the subjects with moderate fine structure. An influence of modulation frequency on MDTs may be predicted by $\mathcal{H}_{sync}$ if one assumes that not only the carrier but also the sidebands may be able to entrain a CR. So when the carrier is at a fine-structure maximum and the sidebands fall at fine-structure minima, the sidebands might entrain CRs, which would increase the effective modulation depth. Note that in this case the detection task would effectively become a discrimination task, but this is not expected to change the thresholds significantly at low modulation depths (Ewert and Dau, 2004). The entrainment would produce a dip in the TMTF of the max condition at $f_{mod}=\Delta f/2$. However, the small peak in the TMTF of the minSL condition cannot be explained by the concept of entrained CRs.

The measured growth functions with tonal carriers are in qualitative agreement with $\mathcal{H}_{sync}$ in that the difference between MDTs at fine-structure minima and maxima disappears with increasing level. In order to test this agreement quantitatively, the individual MDTs at fine-structure minima were simulated in the following way. It was assumed that for every level the MDTs at fine-structure maxima correspond to the internal just noticeable degree of modulation. The MDTs at fine-structure minima were predicted by calculating the external modulation depths that, after the reduction by adding a CR, yielded the same internal modulation depths as in the max condition. The level of the CR was chosen so that the prediction matched the measured data at $L_{car}=15$ dB SL, which is the lowest carrier level at which the data of all subjects are complete. The results of the simulation are shown in Fig. 4 (circles with dashed lines) and the level of the estimated CR is given for every subject. In general, the simulation based on $\mathcal{H}_{sync}$ is able to predict the measured data within ± 2 dB. The deviations in subject SEr are higher, which is probably due to the unusually high MDT at 15 dB SL in the max condition. At low levels many subjects had invalid runs so that the "real" MDT would be expected to be higher than the data point that is shown. In most of these cases the predicted MDT is indeed higher than the measured data and often exceeds 0 dB.

Taking into account the middle-ear transfer function, the levels of SOAEs at f_{car}^{min} can be estimated from the corresponding CR levels. On the one hand the middle ear amplifies the stimulus, which would lead to a higher estimate of L_{CR}; on the other hand the CR is attenuated by the middle ear on its way to the ear canal. Using the middle-ear transfer functions from Puria (2003), SOAE levels in the ear canal should be about 7–22 dB smaller than the predicted CR levels given in Fig. 4, i.e., between -12 and 19 dB SPLs. This is within the range of realistic SOAE levels.

The concept of an addition of carrier and CR in phase as proposed by $\mathcal{H}_{sync}$ may also be tested on the MDT data from experiment 3. The individual MDTs at fine-structure minima

for narrow-band noise carriers were simulated in the same way as described above for tonal carriers. The same CR levels as in the previous simulation were used (see Fig. 4). The results are shown in Fig. 5 as circles with dashed lines. The simulations tend to be slightly higher than the measured data, which indicates that the reduction in effective modulation depth by the CR was actually slightly smaller than assumed by the simulation. This suggests that the CR was not able to fully synchronize to the non-deterministic carrier, which could have two reasons: One possibility is that the instantaneous frequency of the noise carrier may have occasionally dropped outside the frequency region in which the CR could be synchronized. The width of the region of entrainment strongly depends on the level of the external tone relative to the level of the CR; widths between 5 and 20 Hz have been observed by Long and Tubis (1988) for SOAEs. Alternatively, the inherent frequency fluctuations of the narrow-band noise signal may be too fast for the CR to synchronize to. Both concepts lead to a partial lack of synchronization between the carrier at a fine-structure minimum and the CR as proposed by $\mathcal{H}_{\text{sync}}$ and thus provide an explanation for the 3-dB difference observed between the effect sizes for tonal (addition of amplitudes) and narrow-band noise (addition of energy) carriers.

Taken together, the concept of entrained CRs is able to qualitatively and quantitatively explain the level dependence of MDTs at fine-structure minima. Simulations based on $\mathcal{H}_{\text{sync}}$ are in close agreement with the measured data for tonal carriers as well as for narrow-band noise carriers. Most of the SOAEs that could be measured in five subjects (see Fig. 1) can be associated with fine-structure minima, which argues in favor of $\mathcal{H}_{\text{sync}}$. The dependence of MDTs on modulation frequency in regions with strong fine structure, however, cannot be explained by $\mathcal{H}_{\text{sync}}$. An entrainment of CRs by the sidebands of the stimulus could in principle account for part of this dependence in the max condition. However, such an entrainment seems unlikely because it should also lower the MDTs in the max condition below those in the flat condition, which does not agree with the results of experiment 2.

F. The concept of a linear filter ($\mathcal{H}_{\text{factor}}$)

The concept of a frequency-specific gain function—or a linear filter whose frequency characteristics equal the inverted fine structure—implies that the position of the sidebands relative to the fine structure of the threshold in quiet affects modulation perception. More specifically, the modulation-frequency characteristics of MDTs at fine-structure minima and maxima expected under $\mathcal{H}_{\text{factor}}$ resemble the TMTFs measured in the conditions minSL and max in regions with strong fine structure: TMTFs at fine-structure minima have a peak when the sidebands fall at fine-structure maxima and TMTFs at fine-structure maxima have a dip when the sidebands fall at fine-structure minima.

In qualitative agreement with the findings in experiment 3, $\mathcal{H}_{\text{factor}}$ would predict a smaller effect of fine structure on MDTs for narrow-band noise carriers if one assumes a slight reduction in the fine structure due to the finite bandwidth of the carrier (Long and Tubis, 1988).

A linear filter as proposed in $\mathcal{H}_{\text{factor}}$ by definition fails to predict level-dependent MDTs as observed in the present study. A level dependence could be incorporated in $\mathcal{H}_{\text{factor}}$ by assuming a reduction in fine structure with increasing level similar to the reduction in fine structure in equal-loudness contours reported in Mauermann *et al.*, 2004. Still, the filter concept would predict lower MDTs in the max condition than in regions without fine structure, which is seldom observed in the present data. This is especially striking at $f_{\text{mod}}=\Delta f$ where the MDTs from both conditions would be expected to be equal.

Summing up, it seems unlikely that a frequency-specific gain function is responsible for the observed effects of fine structure on MDTs alone. This is consistent with the findings in Heise *et al.*, 2009. However, it appears that in regions with strong fine structure the fine structure does have a filter-like influence on the AM stimuli.

V. SUMMARY AND CONCLUSIONS

The threshold in quiet is commonly used as an indicator for general hearing capabilities. However, a closer inspection reveals that frequency regions with a superior sensitivity to tones do not have a superior sensitivity to amplitude modulations. On the contrary, the modulation sensitivity in fine-structure minima is clearly reduced compared to that in regions without fine structure.

A general increase in MDTs in regions with fine structure may be explained by the beating of the carrier with CRs, which may partly mask the modulation. Such beating, however, cannot account for the large increase in MDTs at fine-structure minima, which was shown by using narrow-band noise carriers (experiment 3). Instead the reduced modulation sensitivity at fine-structure minima may to a large extent be explained by a reduction in effective modulation depth caused by the in-phase addition of a CR to the carrier, which was shown by analyzing MDT growth functions with tonal and narrow-band noise carriers (experiments 2 and 3). In addition, in regions with strong fine structure, fine structure was shown to modify also the modulation-frequency characteristics of MDTs (experiment 1), which agrees with the idea that the cochlear gain is frequency dependent in regions with fine structure.

The findings imply that when measuring modulation sensitivity at levels near the threshold in quiet, care should be taken to avoid artifacts due to threshold fine structure. While for high levels effects of fine structure on modulation perception are expected to be negligible, at low levels the carriers should be placed in regions without fine structure or at least near fine-structure maxima in order to gain comparable results between subjects. The critical level at which the effect of fine structure on MDTs typically disappears can be estimated from the MDT growth functions in experiment 2 at around 30–40 dB SPL [this tallies well with the level at which fine structure tends to disappear from equal-loudness

contours (Mauermann *et al.*, 2004)]. However, this level may vary considerably between subjects depending on their amount of fine structure.

Models that aim at identifying the fundamental mechanisms underlying modulation perception over the whole dynamic range should be able to predict the effect of relatively high modulation thresholds at frequencies with relatively low thresholds in quiet at low stimulus levels.

ACKNOWLEDGMENT

This work was supported by the Deutsche Forschungsgemeinschaft via Grant Nos. GRK591/3 and KO942/18-1&2.

[1]Note that we refer here to evoked OAEs with a single source, which does not include distortion-product OAEs.

[2]In the course of this study a total of 59 ears from 32 subjects (mostly university students) were screened over 1–2 octaves in the range between 1 and 4 kHz. 41 of these ears (i.e., 69%) had fine structure as determined by the FINESS-detector (Heise *et al.*, 2008) with a 3-dB_{avg} criterion (other parameters see below).

[3]The following parameters were used for the FINESS-detector: minimum frequency difference of adjacent fine-structure extreme values $\Delta f_{\min} = 1/50$ octave, maximum frequency differences of adjacent fine-structure extreme values $\Delta f_{\max} \in \{1/8, 1/12, 1/16\}$ octave, minimum level differences of adjacent fine-structure extreme values $\Delta L_{\min} \in \{0, 0.2, \ldots, 10\}$ dB, and the minimum number of extreme values per fine-structure region was three. For the definitions of these parameters and the "dB_{avg}" unit, see Heise *et al.*, 2008.

[4]Note that this quantification of fine structure differs slightly from the quantification that was used above for the general classification of the subjects. Whereas the order of the subjects in the figures is based on the threshold data from all sessions, the correlation analysis only includes data from experiment-specific sessions in order to account for day-to-day variations in the fine structure.

[5]Weber's law is classically known as $\Delta I/I = \text{const}$, describing the ability to detect an intensity increment ΔI at a reference intensity I in a broadband signal. This may be rewritten as $10 \log_{10} \Delta I = 10 \log_{10} I + k$ (where k is a constant), i.e., the slope of $(10 \log_{10} \Delta I)$ versus $(10 \log_{10} I)$ is 1 (e.g., Moore, 2004). AM detection may be regarded as a form of intensity discrimination between the minimum and the maximum intensities of the modulated signal. Then Weber's law may be defined for AM signals by substituting the minimum intensity $(A(1-m))^2$ for I and $(A(1+m))^2 - (A(1-m))^2$ for ΔI.

Bright, K. E. (**1997**). "Spontaneous otoacoustic emissions," in *Otoacoustic Emissions–Clinical Applications*, edited by M. S. Robinette and T. J. Glattke (Thieme, New York).

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

Dau, T., Verhey, J., and Kohlrausch, A. (**1999**). "Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers," J. Acoust. Soc. Am. **106**, 2752–2760.

Elliott, E. (**1958**). "A ripple effect in the audiogram," Nature (London) **181**, 1076.

Ewert, S. D., and Dau, T. (**2004**). "External and internal limitations in amplitude-modulation processing," J. Acoust. Soc. Am. **116**, 478–490.

Glasberg, B. R., and Moore, B. C. J. (**2000**). "Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise," J. Acoust. Soc. Am. **108**, 2318–2328.

Heise, S. J., Mauermann, M., and Verhey, J. L. (**2009**). "Threshold fine structure affects amplitude modulation perception," J. Acoust. Soc. Am. **125**, EL33–EL38.

Heise, S. J., Verhey, J. L., and Mauermann, M. (**2008**). "Automatic screening and detection of threshold fine structure," Int. J. Audiol. **47**, 520–532.

Johnson, T. A., Gorga, M. P., Neely, S. T., Oxenham, A. J., and Shera, C. A. (**2008**). "Relationship between otoacoustic and psychophysical measures of cochlear function," in *Active Processes and Otoacoustic Emissions in Hearing*, edited by G. A. Manley, R. R. Fay, and A. N. Popper (Springer, New York).

Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers," J. Acoust. Soc. Am. **108**, 723–734.

Lee, J., and Bacon, S. P. (**1997**). "Amplitude modulation depth discrimination of a sinusoidal carrier: Effect of stimulus duration," J. Acoust. Soc. Am. **101**, 3688–3693.

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

Long, G. R. (**1984**). "The microstructure of quiet and masked thresholds," Hear. Res. **15**, 73–87.

Long, G. R. (**1993**). "Perceptual consequences of otoacoustic emissions," in *Contributions to Psychological Acoustics: Results of the 6th Oldenburg Symposium on Psychological Acoustics*, edited by A. Schick (University of Oldenburg Press, Oldenburg, Germany).

Long, G. R., and Tubis, A. (**1988**). "Investigations into the nature of the association between threshold microstructure and otoacoustic emissions," Hear. Res. **36**, 125–138.

Mauermann, M., Long, G. R., and Kollmeier, B. (**2004**). "Fine structure of hearing threshold and loudness perception," J. Acoust. Soc. Am. **116**, 1066–1080.

McFadden, D., and Mishra, R. (**1993**). "On the relation between hearing sensitivity and otoacoustic emissions," Hear. Res. **71**, 208–213.

Millman, R. E., and Bacon, S. P. (**2008**). "The influence of spread of excitation on the detection of amplitude modulation imposed on sinusoidal carriers at high levels," J. Acoust. Soc. Am. **123**, 1008–1016.

Moore, B. C. J. (**2004**). *An Introduction to the Psychology of Hearing* (Elsevier Academic, London).

Moore, B. C. J., and Glasberg, B. R. (**2001**). "Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 1067–1073.

Puria, S. (**2003**). "Measurements of human middle ear forward and reverse acoustics: Implications for otoacoustic emissions," J. Acoust. Soc. Am. **113**, 2773–2789.

Shera, C. A., and Zweig, G. (**1993**). "Order from chaos: Resolving the paradox of periodicity in evoked otoacoustic emission," 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).

Strickland, E. A., and Viemeister, N. F. (**1997**). "The effects of frequency region and bandwidth on the temporal modulation transfer function," J. Acoust. Soc. Am. **102**, 1799–1810.

Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). "Modeling otoacoustic emission and hearing threshold fine structures," J. Acoust. Soc. Am. **104**, 1517–1543.

Wojtczak, M., and Viemeister, N. F. (**2008**). "Perception of suprathreshold amplitude modulation and intensity increments: Weber's law revisited," J. Acoust. Soc. Am. **123**, 2220–2236.

Yost, W. A., and Sheft, S. (**1997**). "Temporal modulation transfer functions for tonal stimuli: Gated versus continuous conditions," Aud. Neurosci. **3**, 401–414.

Zweig, G., and Shera, C. A. (**1995**). "The origin of periodicity in the spectrum of evoked otoacoustic emissions," J. Acoust. Soc. Am. **98**, 2018–2047.

Zwicker, E. (**1986**). "Spontaneous oto-acoustic emissions, threshold in quiet, and just noticeable amplitude modulation at low levels," in *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson (Plenum, New York).

Zwicker, E., and Schloth, E. (**1984**). "Interrelation of different oto-acoustic emissions," J. Acoust. Soc. Am. **75**, 1148–1154.

Level dependence in behavioral measurements of auditory-filter phase characteristics

Yi Shen[a] and Jennifer J. Lentz
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405-7000

(Received 31 October 2008; revised 11 August 2009; accepted 13 August 2009)

Two masking experiments were conducted to behaviorally estimate auditory-filter phase curvatures at different stimulus levels. Maskers were harmonic complexes consisting of equal-amplitude tones and phase spectra with varied curvatures. In Experiment 1, sinusoidal signal thresholds were measured at 2 and 4 kHz at fixed masker levels ranging from 50 to 90 dB sound pressure level (SPL). In Experiment 2, the masker level that just masked a sinusoidal signal at 2 and 4 kHz was measured at fixed signal levels of 25, 38, and 50 dB SPL. For both experiments, the estimated phase curvature approached zero (became less negative) with increasing stimulus level. This shift could suggest that the off-frequency phase characteristic of the auditory filter has an increasingly greater role on the estimated auditory-filter phase curvature at higher stimulus levels. This explanation is supported through the use of psychophysical modeling.
 [DOI: 10.1121/1.3224709]

PACS number(s): 43.66.Nm, 43.66.Dc, 43.66.Ba [BCM] Pages: 2501–2510

I. INTRODUCTION

Historically, masking experiments have been successful tools in describing the frequency selectivity of the auditory system (e.g., Fletcher, 1940) and the magnitude response of auditory filters (Patterson *et al.*, 1982), but more recently, there also has been a focus on investigating the phase response of the auditory system using behavioral methods (e.g., Lentz and Leek, 2001; Oxenham and Dau, 2001b). These experiments employ harmonic complexes as maskers rather than the noise stimuli used in many classic masking experiments. Smith *et al.* (1986) and Kohlrausch and Sander (1995) were among the first to introduce the idea of estimating the phase characteristics of the auditory system using harmonic stimuli. Both studies showed that the phase relationships of the components of a harmonic masker can greatly influence the detection threshold of a pure-tone signal, even for stimuli with identical power spectra. Two specific phase settings in the maskers, namely, the positive Schroeder phase (+Schr) and the negative Schroeder phase (−Schr) [described by Schroeder (1970)], can lead to threshold differences as large as 20 dB. The +Schr stimulus (characterized by a downward linear frequency sweep) is the time reverse of −Schr stimulus and is typically a less effective masker than −Schr stimulus. This difference in thresholds for stimuli with identical power spectra but different phase spectra is called the masker phase effect.

It is thought that the masker phase effect arises from interactions between the frequency glide of the auditory-filter impulse response and the frequency glides of the complex maskers (Kohlrausch and Sander, 1995). In particular, the impulse response of the auditory filter has a low-to-high frequency glide. The rate of the glide, which is related to the phase curvature, interacts with that of the masker. Kohlrausch and Sander (1995) argued that when the phase curvature of the cochlea counteracts that of a masker, the internal representation of the masker has all frequency components in phase. As a consequence, a highly modulated temporal envelope is formed. When a tone is added to this internally modulated masker, the resulting internal representation will contain valleys that have a higher signal-to-noise ratio than the peaks. The higher signal-to-noise ratio in the valleys provides a better chance to detect a target signal and could lead to a lower detection threshold than for a less-peaked waveform having no distinct peaks and valleys. Psychophysical masking period patterns (MPPs) and physiological experiments support this interpretation by demonstrating that the +Schr stimulus has a more peaked representation than the −Schr stimulus (Kohlrausch and Sander, 1995; Recio and Rhode, 2000; Summers *et al.*, 2003).

Closer evaluation of the relationships between masker phase and masked threshold can be made using a procedure developed by Lentz and Leek (2001) by introducing a scalar C into the original Schroeder-phase formula:

$$\theta_n = C\pi n(n-1)/N, \tag{1}$$

where θ_n is the phase of the nth harmonic and N is the total number of components. Scalar C is proportional to the phase curvature and hence is inversely related to the rate of the frequency glide in the complex. This modification allows a systematic manipulation of the phase settings ranging between negative Schroeder phase ($C=-1$) and positive Schroeder phase ($C=1$) without altering the spectral composition of the stimulus. Using this equation, the C value at which the masking efficiency reaches a minimum ($C_{\min}$) can be measured. This $C_{\min}$ value corresponds to the phase setting in the masker that best mirrors the cochlear phase response and therefore provides an indirect measure of the auditory-filter phase curvature. $C_{\min}$ is not necessarily the positive Schroeder-phase setting, but all estimates of $C_{\min}$ in

[a]Author to whom correspondence should be addressed. Electronic address: shen2@indiana.edu

humans have been positive in sign (Lentz and Leek, 2001; Oxenham and Dau, 2001b, 2004).

Although the cochlear nonlinearity has been implicated in the size of the masker phase effect, little psychophysical work exists which assesses whether the behaviorally estimated phase curvature is level dependent. Most studies have evaluated level dependence in Schroeder-phase masking using only +Schr and −Schr stimuli as maskers. Carlyon and Datta (1997a) showed that for masking of long-duration tones, the difference in masking efficiency between +Schr and −Schr stimuli tends to increase with increasing stimulus level. Carlyon and Datta (1997b) also indicated that MPPs of −Schr stimuli do not typically change with level, but the degree to which the MPPs for +Schr stimuli vary with time tends to increase with increasing stimulus level and decreases at very high levels (see also, Summers, 2000).

This non-monotonic dependence on level can be understood as a trade-off between two separate effects. On the one hand, increases in level cause modulated sounds to become less effective as maskers than unmodulated sounds (Bacon *et al.*, 1997). Because the +Schr stimulus has a greater internal modulation than the −Schr stimulus, it would become a (relatively) less effective masker as stimulus level increases. Consistent with this idea is the finding that the difference in masked thresholds provided by the +Schr and −Schr stimuli tends to increase with increasing stimulus level (Carlyon and Datta, 1997a).

On the other hand, level increases cause the temporal modulation of +Schr and −Schr stimuli to become more similar. This effect has been demonstrated physiologically by Summers *et al.* (2003), who measured basilar-membrane (BM) responses to Schroeder-phase complexes in guinea pig using laser velocimetry. Their results show that while the envelopes of BM responses to −Schr stimuli do not vary greatly with level, the BM response envelopes of +Schr stimuli show a reduction in temporal modulation depth as level increases. At low levels, the +Schr BM responses are significantly peakier than those of −Schr stimuli and are similar to BM impulse responses derived using broadband noises. In contrast, at high levels, the difference in modulation depth between +Schr and −Schr stimuli is much less obvious. Summers *et al.* (2003) argued that these results could have arisen due to an auditory-filter phase curvature that is not constant along the frequency axis. This phase curvature has a large negative curvature at the characteristic frequency (CF) and a reduced curvature (nearer to zero) at lower frequencies. At low stimulus levels, then, a single location along the BM responds to a narrow frequency region around the CF, and a positive Schroeder-phase curvature could compensate the curvature of the auditory-filter phase response thereby leading to a peaked BM response. As level increases, however, the response region of this same location extends to lower frequencies where the +Schr setting does not mirror the auditory-filter phase response. As a result, the frequency components in +Schr stimuli do not arrive at the observation location synchronously, resulting in a less-peaked response envelope.

Predictions can be drawn from this "off-frequency phase influence" hypothesis about the behaviorally estimated auditory-filter phase curvature. Specifically, when measuring signal detection thresholds as a function of masker phase curvature (by systematically varying C), it is expected that $C_{\min}$ would be more positive at lower levels than at higher levels. To our knowledge, Oxenham and Dau (2001b) provided the only experiment that systematically explored whether changes in stimulus level lead to changes in the behaviorally estimated auditory-filter phase curvature. Three masker levels (40, 60, and 85 dB) were tested at each of three signal frequencies (250, 1000, and 4000 Hz). They found that $C_{\min}$ did not change significantly with stimulus level, which was in contradiction with the hypothesis proposed by Summers *et al.* (2003). The present study uses a pair of experiments to expand upon the work of Oxenham and Dau (2001b). These experiments are intended to determine whether the estimate of auditory-filter phase curvature varies with stimulus level by using a larger sample of masker levels and by testing different C values at various masker and signal levels.

II. EXPERIMENT 1: EFFECTS OF MASKER LEVEL ON ESTIMATED PHASE CURVATURE

A. Methods

1. Stimuli

Thresholds were measured for a sinusoidal signal in the presence of a simultaneous masker. Both the signal and the masker were 300 ms in duration. They were gated together with 30-ms raised-cosine onset and offset ramps. Two signal frequencies, f_s, were tested: 2000 and 4000 Hz. These signal frequencies were selected to improve chances of pinpointing minima in the functions, which are expected to be $C=1.0$ and below for the chosen masker fundamental frequency and masker bandwidth in the present experiment.[1] On every stimulus presentation, the starting phase of the signal was selected randomly from a distribution of $0-2\pi$ radians. The random starting phase of the signal was chosen so that our thresholds can be compared to those of other studies in which scalar C was varied (e.g., Oxenham and Dau, 2001b).

The masker was a harmonic tone complex with a fundamental frequency of 100 Hz and frequency components ranging between $0.4f_s$ and $1.6f_s$. The phases of the components were selected according to a modification of Schroeder's phase equation (Lentz and Leek, 2001). For each signal frequency, the masker was presented at fixed overall levels of 50-, 60-, 70-, 80-, and 90-dB sound pressure level (SPL). These levels are about 14 dB higher than the masker component level for $f_s=2$ kHz and 17 dB higher for $f_s=4$ kHz. An additional high-pass broadband noise (cutoff frequency $=1.8f_s$) was presented at a total power of 50-dB SPL simultaneous with the complex masker in order to limit off-frequency listening to frequencies beyond that of the masker. Although this masker may not have been at a high enough level to limit off-frequency listening, spot checks on two of the subjects with a high-pass masker presented at 80-dB SPL indicated no difference in threshold values for the 50- and 80-dB SPL masker levels. Thus, it is expected that listeners were not listening off frequency. For each f_s and stimulus level, signal detection thresholds were measured for C values

ranging from −1.0 to +1.5. At least seven C values were tested at each masker level; these C values were chosen based on pilot data and varied at the different masker levels.

The stimuli were generated digitally and presented using a 24-bit Tucker-Davis-Technologies Real-Time processor (TDT RP2.1; sampling rate=48 828 Hz), a programmable attenuator (TDT PA5), and a headphone buffer (TDT HB6). Stimuli were then presented monaurally via Sennheiser HD250 II Linear headphones. The experiment was conducted in a double-walled sound-attenuating booth.

2. Procedure

An adaptive three-interval, three-alternative forced-choice procedure was used in conjunction with a 2-down, 1-up tracking rule to estimate the 70.7%-correct point on the psychometric function (Levitt, 1971). The masker stimulus was presented in all three intervals, with the signal stimulus being added to any one of the three intervals with equal probability. Within each trial, the three intervals were indicated by LED lights and separated by 500-ms silent pauses. The participants responded to the stimuli via a button box and were given correct-answer feedback through the LED lights. Each track consisted of eight reversals. The track began with a signal level that was equal to that of the masker. The initial step size was 5 dB, which was reduced to 2 dB after the first two reversals. Threshold was defined as the mean of the signal levels at the final six reversals.

The experiment was divided into two separate sections with all listeners being tested at 2000 Hz before being tested at 4000 Hz. For each signal frequency, the presentation order of the masker levels was randomly chosen with the constraint that the measurements at high masker levels (80- and 90-dB SPL) were not run in adjacent blocks. Once the masker level was selected, all the C values were tested in random order before the next masker level was chosen. After all masker levels were tested once for a given signal frequency, a new random order of masker levels was selected and the process repeated. Thresholds were measured at least four times at each masker level for each signal frequency. A final threshold was based on the average of these four threshold measurements. When the standard deviation across these four threshold estimates exceeded 8 dB, two more threshold estimates were included in the mean threshold. This happened for only two threshold estimates for one of the subjects (NH3). Measurements were conducted in 1.5-h sessions spanning seven to eight visits with no more than one session per day for each subject.

3. Subjects

Four subjects (two male) participated. One was the first author (NH4), and the other three were paid on an hourly basis. The subjects' ages ranged from 24 to 28 years. The pure tone thresholds of all subjects were 10-dB hearing level or better for audiometric frequencies between 250 and 6000 Hz, and the ear with better audiometric thresholds was tested. Subjects NH1 and NH3 had no previous experience in psychoacoustic experiments and received about 1 h of training before data collection started.

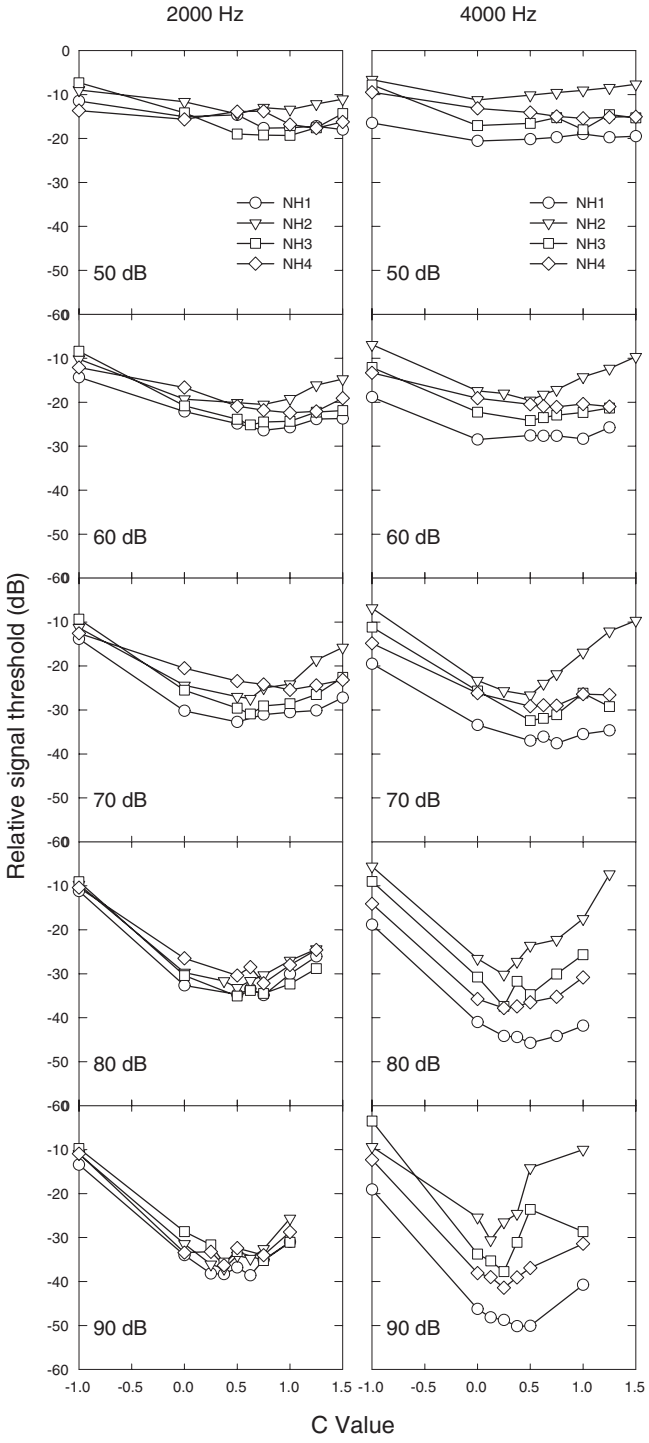


FIG. 1. Individual masked thresholds, relative to the overall masker level, are plotted as a function of scalar C, for f_s=2000 Hz (left) and 4000 Hz (right). Results for overall masker levels of 50-, 60-, 70-, 80-, and 90-dB SPL are shown in separate panels.

B. Results

The individual data for f_s=2000 and 4000 Hz are shown in the left and right panels of Fig. 1, respectively. Masked thresholds, expressed as the signal level relative to the overall masker level in decibels, are plotted as a function of the scalar C. Results for masker levels of 50-, 60-, 70-, 80-, and 90-dB SPL are shown in separate panels, and different symbols indicate data obtained from the four individual listeners.

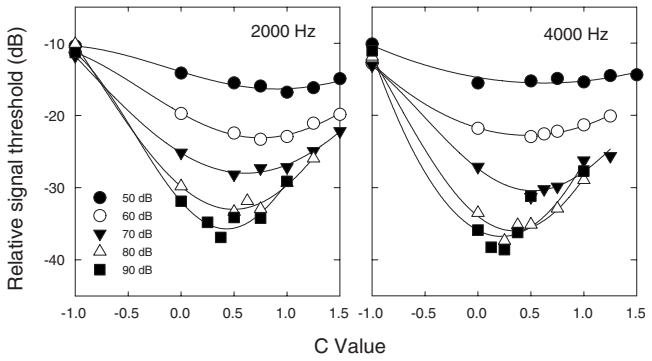


FIG. 2. Mean masked thresholds across subjects, relative to the overall masker level, are plotted for f_s=2000 Hz (left) and 4000 Hz (right). Results for the five overall masker levels are denoted with different symbols. Solid lines indicate data fits with a sinusoidal function $[y=y_0+a\sin((2\pi x/b)+c)]$, with y_0, a, b, and c as free parameters.

For both signal frequencies, all individuals show a similar pattern in the functions relating masked threshold to C. In general, as scalar C increases, threshold decreases and then increases forming a dip centered around a positive C value. At 2000 Hz, standard errors across the four listeners are typically below 2.5 dB. Greater variability in masked thresholds is present across listeners at 4000 Hz (standard errors increase with increasing stimulus level and reach 7-dB at 90-dB SPL). Although Fig. 1 reveals large variability across listeners, such variability is commonly reported in masking studies using Schroeder-phase maskers. For very similar stimuli, the data of Oxenham and Dau (2001b) have standard errors of 5 or 6 dB (across four listeners) at 4000 Hz (see error bars in Fig. 6 of their paper).

The mean experimental data are shown in Fig. 2, which plots thresholds obtained at different masker levels as different symbols. At each masker level, the average threshold as a function of scalar C resembles the individual data, where the threshold curves form dips around positive C values. We will call the C value where the threshold reaches its minimum $C_{\min}$ and name the threshold difference between $C=-1$ and $C=C_{\min}$ the depth of threshold. Across different stimulus levels, we see that although the thresholds at $C=-1$ ($-$Schr) are quite stable, the thresholds at $C=C_{\min}$ vary greatly with level. As a consequence, the depth of threshold increases as the masker level increases.

The value of $C_{\min}$ tends to shift toward lower C values with increases in the masker level. $C_{\min}$ was estimated using one of the methods suggested by Oxenham and Dau (2001b). The mean data in each condition were fitted with a sinusoidal function, $y=y_0+a\sin((2\pi x/b)+c)$, where y_0, a, b, and c are four free parameters. The best-fitting functions (in a least-squares sense) are plotted in Fig. 2 as the solid lines. The minima of these functions at the different masker levels are taken as estimated values of $C_{\min}$. For f_s=2000 Hz, the minimum tends to be near 0.8 at 60-dB SPL and shifts to about 0.4 at 90-dB SPL. For f_s=4000 Hz, the minimum is near 0.5 at 60-dB SPL and is between 0.2 and 0.3 at 90-dB SPL. For both signal frequencies, there is a change of a factor of 2 in the minimum between 60- and 90-dB SPL. Although there is some variability in $C_{\min}$ across listeners (note Fig. 1), all listeners show decreasing shifts in $C_{\min}$ with increasing stimulus level.

The thresholds reported in Fig. 2 obtained at 4000 Hz can be compared directly with those reported by Oxenham and Dau (2001b) in their Fig. 6, as we used the same fundamental frequency and range of frequencies for the 4000-Hz signal. Our average thresholds are approximately 15 dB higher than those in Oxenham and Dau (2001b). However, Oxenham and Dau pointed out an error in their figure, which is that the thresholds are actually plotted relative to overall masker level and not (as is stated in the caption of their Fig. 6) relative to masker component level (Oxenham and Dau, personal communication). Taking this correction into account, the thresholds of Fig. 2 are quite similar to those of Oxenham and Dau (2001b). Oxenham and Dau (2001b) reported 4000-Hz thresholds in terms of signal-to-noise ratio between -10 and -15 dB at $C=-1$ for all stimulus levels. Our thresholds for this same condition range between -10 and -13 dB. Thresholds are also similar between the two studies for C values greater than zero. For example, data of Oxenham and Dau (2001b) at 85-dB SPL are between -30 and -35 dB for $C\geq 0$, and Fig. 1 shows threshold levels between about -30 and -40 dB at 80 and 90-dB SPL.

The biggest discrepancy between our data and those of Oxenham and Dau (2001b) is that our data show decreases in $C_{\min}$ with increases in stimulus level whereas the data of Oxenham and Dau (2001b) do not. It is not obvious what causes this discrepancy between the two experiments. Some possible reasons are as follows: Oxenham and Dau (2001b) tested the 4000-Hz signal in the presence of a low-pass masker to limit the effect of low-frequency distortion at high stimulus levels, whereas the current experiment did not include a low-frequency noise masker. To assess whether the presence of the low-pass masker might have influenced the pattern of results, thresholds for the signal frequency of 2000 Hz were measured again for NH4 at two masker levels (60- and 90-dB SPL) in the presence of a low-pass masker.[2] These thresholds did not differ from the previous thresholds, and the $C_{\min}$ did not change. This result suggests the shift of $C_{\min}$ in our experiment is not a consequence of low-frequency distortion at high stimulus levels. It is also possible that the present experiment had a better chance of detecting level dependence because the C values were chosen specifically to optimize $C_{\min}$ estimates and the masker level was altered in 10-dB steps. Oxenham and Dau (2001b) only tested $C\leq 1.0$ and may not have been able to pinpoint a minimum in some of the functions. Note that in Fig. 2, some of the functions, especially those at low levels, reveal minima only because C values greater than 1.0 were tested.

C. Estimating the phase curvature of the auditory filters

In this experimental paradigm, lower masked thresholds are thought to reflect more peaked internal waveforms, providing listeners the opportunity to "listen-in-the-valleys" (Buus, 1985) more than when internal waveforms are not as peaked. Because it has been argued that these highly peaked waveforms reflect the greatest interaction between the

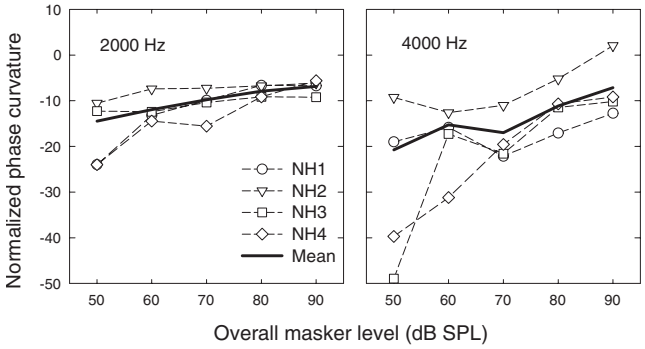


FIG. 3. The estimated phase curvature for f_s=2000 (left) and 4000 Hz (right), transformed into dimensionless units by multiplying by $f_s^2/2\pi$. The open symbols and the solid lines represent the estimated phase curvature based on curve fits to the individual data and the mean data, respectively.

auditory-filter phase curvature and the cochlear phase curvature, the C value that provides the lowest detection threshold ($C_{\min}$) might be considered an indirect measure of the auditory-filter phase curvature (Kohlrausch and Sander, 1995). Estimates of the phase curvature are based on an assumption that the phase curvature within the auditory-filter passband is constant and has the same magnitude and opposite sign to the phase curvature of the complex masker.

Using the estimates of $C_{\min}$ from the mean data, the phase curvature of the auditory filter was calculated from

$$\frac{d^2\theta}{df^2} = -C_{\min}\frac{2\pi}{Nf_0^2}, \tag{2}$$

where N is the total number of components and f_0 is the fundamental frequency of the masker. This procedure also was carried out on the individual data to assess whether individual differences in the phase curvature as a function of the masker level are present. $C_{\min}$ was estimated by fitting a sinusoidal function to the individual data in the same manner as was done for the mean data. It should be noted that at lower masker levels, the threshold varies little with the scalar C, and therefore the estimate of the phase curvature has greater error.

For ease in comparing across frequencies, absolute phase curvature values were normalized by multiplying the estimated curvature by $f_s^2/2\pi$ (Shera, 2001). The resulting quantities are dimensionless. Figure 3 shows the normalized phase curvature as a function of overall masker level for f_s =2000 and 4000 Hz.

The magnitude of the estimated auditory-filter phase curvature based on the mean data tends to decrease with increasing masker level (i.e., it approaches zero) for both signal frequencies. The estimated curvature also plateaus above 70-dB SPL at 2000 Hz for three of the four listeners, but no plateau region is observable at 4000 Hz. At low masker levels, the curvature at 4000 Hz appears to be more negative than at 2000 Hz. However, there is a great deal of variability across individual subjects, and the curvature does not differ markedly for the two signal frequencies. These curvature estimates agree well with the results from other studies where auditory-filter phase curvatures were estimated either at a fixed masker level (Oxenham and Dau, 2001b) or for a fixed signal level (Lentz and Leek, 2001). Using a 75-dB SPL masker, Oxenham and Dau (2001b) reported normalized auditory-filter phase curvature estimates of about −10 and −17 for 2000 and 4000 Hz, respectively. Lentz and Leek (2001) used a fixed-signal level of 40-dB SPL and found curvatures of about −8 and −25 at 2000 and 4000 Hz. These previous studies provide a range of curvature estimates that are consistent with our average estimates of −10 and −16 at 2000 and 4000 Hz for a 70-dB SPL masker.

To confirm that stimulus level influences the estimated auditory-filter phase curvature, a repeated-measures analysis of variance treating masker level and signal frequency as within-subject factors revealed a significant effect of level [$F(4,12)=10.99$, $p<0.005$] but not signal frequency [$F(1,3)=6.02$, $p=0.09$]. There was no significant interaction between masker level and signal frequency [$F(4,12)=0.80$, $p=0.55$], suggesting that the changes in phase curvature with stimulus level do not vary across these frequencies. Figure 3 also shows individual differences in the rate of change of curvature with masker level. For example, at 2000 Hz, NH1 and NH4 show larger shifts in curvature with masker level than NH2 and NH3. At 4000 Hz, NH1 shows a smaller shift of curvature than the other subjects. Despite these different rates, the estimated magnitude of the auditory-filter phase curvature for all subjects decreases with increasing masker level.

One must be cautious, however, in interpreting these results as providing support for level-dependent changes in the phase characteristic of the auditory filter. First, this observation is based on the estimation of $C_{\min}$ and an assumption that the phase curvature is constant across the auditory-filter passband. Pinpointing $C_{\min}$ can be difficult due to a clustering of low threshold values near $C_{\min}$ (see Figs. 1 and 2, especially at lower masker levels), and the assumption of constant phase curvature may not hold. Second, if the off-frequency phase response of the auditory filter significantly influences the response of the BM to the complex masker as suggested by Summers *et al.* (2003), a level-dependent shift in $C_{\min}$ is expected regardless of whether the auditory-filter curvature at the CF is varying with level. Finally, the level-dependent shifts in curvature reported here do not replicate the findings of Oxenham and Dau (2001b) who did not observe a level-dependent shift in $C_{\min}$ even though they used a very similar experimental design. Given that the source of the difference in experimental results is unknown, Experiment 2 tests whether the results of Experiment 1 generalize using a different experimental paradigm.

III. EXPERIMENT 2: EFFECTS OF SIGNAL LEVEL ON ESTIMATED PHASE CURVATURE

A. Methods

The parameters of the harmonic complex stimuli used in the present experiment were identical to those used in the first experiment. The overall masker level that just masked a signal at different signal levels was measured for C values ranging from −1.0 to 1.5. The signal levels were 25-, 38-, and 50-dB SPL for f_s=2000 Hz, and were 25- and 38-dB SPL for f_s=4000 Hz. Based on pilot listening, the 50-dB

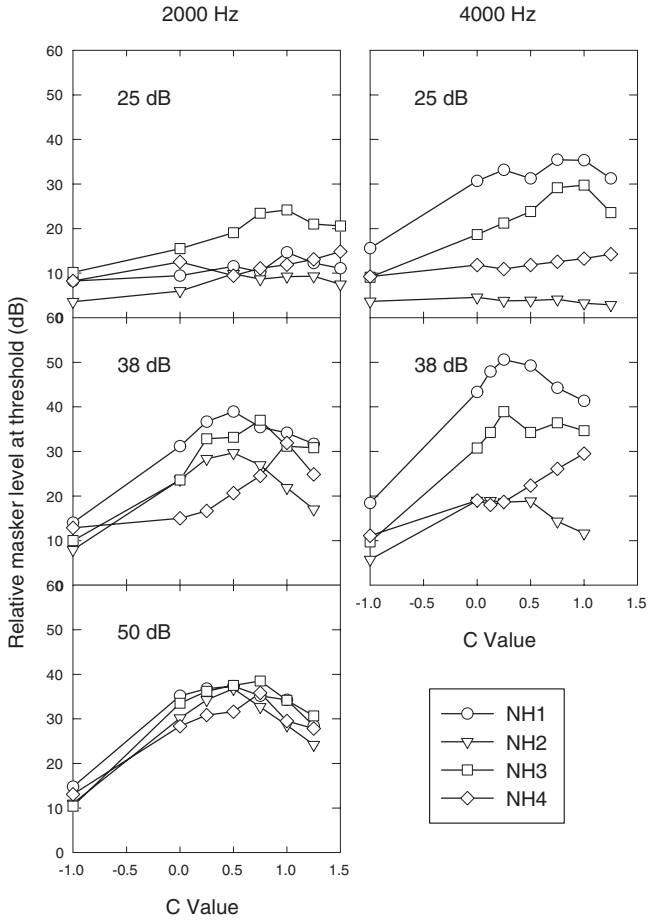


FIG. 4. Individual overall masker levels at threshold, relative to the signal level, are plotted as a function of C for f_s=2000 Hz (left) and 4000 Hz (right). Results from the measurements for signal levels of 25-, 38-, and 50-dB SPL are shown in separate panels.

SPL signal level was excluded from the 4000-Hz condition to prevent presenting excessively high masker levels (greater than 100-dB SPL). The lowest signal level, 25-dB SPL, was at least 10 dB above the absolute threshold for all subjects. The methods for estimating the masker level needed just to mask the signal were the same as in the first experiment, with the following exceptions. The overall masker level was increased after two consecutive correct responses and decreased after one incorrect response, and the total number of reversals for each track was 10. The initial step size was 8 dB, which was reduced to 5 dB after the first two reversals, and reduced to 2 dB after another two reversals. The estimated threshold was the mean of the masker levels at the final six reversals. A final threshold was based on the average of 4 repetitions. All four subjects from the first experiment participated and the same ears were tested.

B. Results

The masker levels that just masked the signal are plotted in Fig. 4 as a function of C. The individual data for f_s =2000 and 4000 Hz are shown in the left and right panels, respectively. Lower masker levels at threshold indicate more effective maskers. The trends observed in this experiment are generally similar to those found in Experiment 1. First, as the scalar C increases, the masker level usually reaches a peak value at a positive C value. However, for subject NH4, this maximum cannot be observed for the function relating masker level to C at low signal levels for both signal frequencies. Second, masker levels at threshold are typically lowest at $C=-1$, indicating that the −Schr stimulus is the most effective masker. Third, the relative masker levels at threshold are generally higher at higher signal levels, suggesting that, relatively speaking, the higher-level maskers are not as effective as the lower-level maskers.

The differences in masked threshold across individuals tend to be much larger than those reported in Experiment 1 and are again greater at 4000 Hz than at 2000 Hz. These individual differences typically occur across all C values tested and are consistent with the individual variability present in Experiment 1. For example, Fig. 4 shows that at 4000 Hz, NH1 had the highest masked thresholds, reflecting little susceptibility to masking, whereas NH2 had much lower thresholds, reflecting greater susceptibility to masking. These two listeners also experienced the most (NH2) and least (NH1) masking in Experiment 1. This across-observer variability is not likely to be due to variability within a subject because for each subject, the standard deviation of the four repetitions revealed relatively small within-subject variability. The similarity across experiments suggests that intrinsic observer-dependent sources are responsible for the large across-observer variability. Large individual differences have also been observed by Lentz and Leek (2001) who measured the levels of Schroeder-phase stimuli required to mask a 40-dB SPL tone at 2000 and 4000 Hz. At 2000 Hz, one subject had thresholds that were consistently 20–25 dB higher than the other subjects, and at 4000 Hz, a different subject had thresholds that were 10–15 dB higher than the others.

As in Experiment 1, the points of minimum masking ($C_{\min}$) corresponding to the peak masker levels were estimated in order to investigate the level dependence of the phase curvature of the auditory filter. Due to the large individual differences, this analysis was based on the individual data only. The same procedure described for Experiment 1 was performed to estimate $C_{\min}$. Briefly, results were fitted with a sinusoidal function. The C at the maximum of the function is estimated as $C_{\min}$. The conditions in which the fitted function failed to reach a maximum in the range of $-1<C<1.5$ was excluded from further analysis. The phase curvatures derived from these estimated $C_{\min}$ are shown in Fig. 5 as functions of signal level for f_s=2000 and 4000 Hz. A repeated-measures analysis of variance was conducted based on the relatively complete estimation data from NH1, NH2, and NH3 at 2000 Hz, treating signal level as a within-subject factor. The analysis revealed a significant effect of signal level on the estimated phase curvatures [$F(2,4)$ =37.16, p=0.003]. This demonstrates that the psychophysically estimated phase curvature shifts toward zero as signal level increases, and replicates the results of Experiment 1 using a different paradigm.

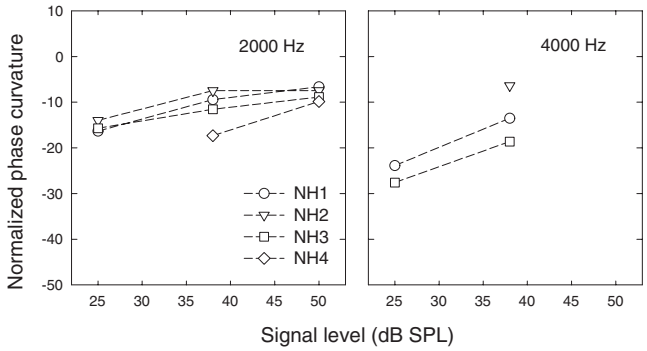


FIG. 5. The estimated phase curvature for f_s=2000 (left) and 4000 Hz (right), transformed into dimensionless units by multiplying by $f_s^2/2\pi$. The estimated phase curvature was based on curve fits to the individual data in Experiment 2.

IV. DISCUSSION

A. Relationship to physiological findings

In the present study, we investigated whether the behaviorally estimated auditory-filter phase curvature was level dependent using a Schroeder-phase masking paradigm. Two experiments revealed a decrease in the magnitude of the phase curvature of the least-efficient masker with increasing stimulus level. One possible interpretation of these results could be that the auditory-filter phase curvature is level dependent. However, this interpretation contrasts physiological measurements of BM vibration (de Boer and Nuttall, 1997; Recio *et al.*, 1998) and auditory nerve response (Carney *et al.*, 1999) in which level-invariant frequency glides have been observed in the impulse responses. Given the contradiction, we must consider the possibility that the phase curvature estimated from the psychoacoustic experiments is not the phase curvature of the auditory filter *per se*. It is likely to be influenced by the near-zero phase curvature in the frequency region away from the CF (Summers *et al.*, 2003).

The auditory-filter curvature directly relates to the instantaneous frequency (IF) trajectory of the auditory-filter impulse response. Specifically, the IF trajectory is the inverse function of the filter's group delay (e.g., Shera, 2001), and the phase curvature is defined as the (negated) slope of the group delay. Because the IF trajectory of the auditory-filter impulse response typically exhibits a low-to-high frequency glide due to the dispersion of the cochlear traveling wave, the auditory-filter curvature is almost always negative in sign. The rate of the glide in the IF trajectory varies greatly as the IF approaches the CF of the filter, causing the auditory-filter curvature to be frequency dependent. This implies that behavioral measurements of the curvature using Schroeder complexes, which assume that the auditory-filter curvature is roughly constant within the passband of the auditory filter, might be based on an invalid assumption.

If the Schroeder-phase masking experiment "pinpoints" the auditory-filter phase curvature of a specific frequency, our experimental results would directly suggest a level-dependent auditory-filter phase curvature. The discrepancy between the behavioral data and the physiological findings would be unresolvable. Further studies would be needed to identify whether the slight changes in the auditory-filter phase curvature that were undetectable physiologically could give rise to significant perceptual consequences. On the other hand, if the "off-frequency phase influence" hypothesis (Summers *et al.*, 2003) is correct, behavioral curvature estimation would be based on the integration of the auditory-filter curvature over frequency. In this case, one might measure a level-dependent psychophysical curvature estimate even though a level-invariant auditory-filter curvature is present. This will be described below.

The behavioral curvature estimation could reflect an interaction between the level-dependent magnitude response of the auditory filter and the distribution of the auditory-filter phase curvature along the frequency axis. At low levels, the shape of the auditory filter is relatively narrow, and the frequency components nearest the signal frequency provide the most masking. These components fall into the portion of the phase response that changes the most with frequency and has a negative phase curvature. In contrast at high levels, the low-frequency skirt of the auditory filter tends to broaden leading to a greater masking contribution from low-frequency components than at lower stimulus levels. The filter's phase response to these low-frequency components is linear and has a curvature near zero. If those low-frequency components have a large influence on the response at the output of the filter, the estimated phase curvature would reflect a different phase curvature than the curvature near the center of the filter (Oxenham and Ewert, 2005). In this way, the behavioral curvature estimates could shift with level even if the filter has a level-invariant auditory-filter curvature.

In terms of the impulse response of the auditory filter, the hypothesis can also be described in the time domain. As stimulus level increases, the envelope of the impulse response becomes increasingly asymmetric, with an increased emphasis on the earlier portion of the impulse response where the IF is lower and the slope of the frequency glide is steeper. As a consequence, the estimated psychophysical curvature has a magnitude that decreases with increasing stimulus level.

B. Model demonstration

To demonstrate the viability of the hypothesis that the estimation of the behavioral curvature reflects an interaction between a level-dependent magnitude response and a level-invariant phase response, model predictions of two artificial auditory-filter models (Models A and B) are compared. Model A is an auditory filter with a level-varying magnitude response but a constant curvature in the phase response (as typically assumed by psychophysical estimates of Schroeder-phase masking). Model B has the same magnitude response as Model A, but a frequency-variant phase curvature. In both filter models, the magnitude responses are identical to those of the corresponding level-dependent gammachirp filter (Irino and Patterson, 2001), but the phase responses are forced to be level invariant. Because the phase responses of the filters are artificially manipulated, these test filters are not intended to reflect cochlear processing, but are being used to demonstrate that a level-invariant auditory-filter phase response can lead to level-varying estimates of the phase cur-

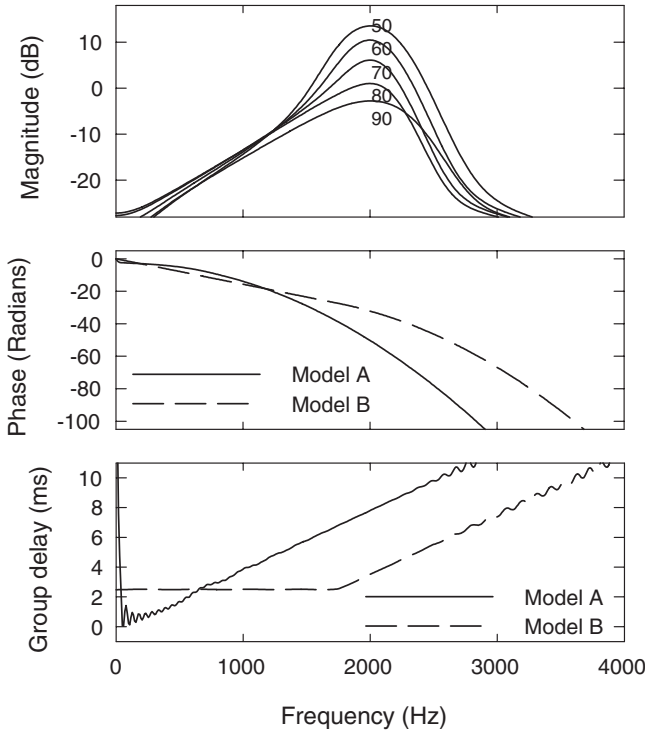


FIG. 6. Magnitude (top), phase (mid), and group delay (bottom) of two artificial auditory-filter models, Models A and B. These two filters have the same magnitude response adopted from a level-dependent gammachirp filter centered at 2000 Hz, illustrated in the top panel for the five masker levels tested in Experiment 1. A single level-invariant phase response (and therefore group delay) of the filters is used for each of the stimulus levels, with Model A having a constant curvature and Model B having a curvature of zero below 1750 Hz and a non-zero curvature elsewhere.

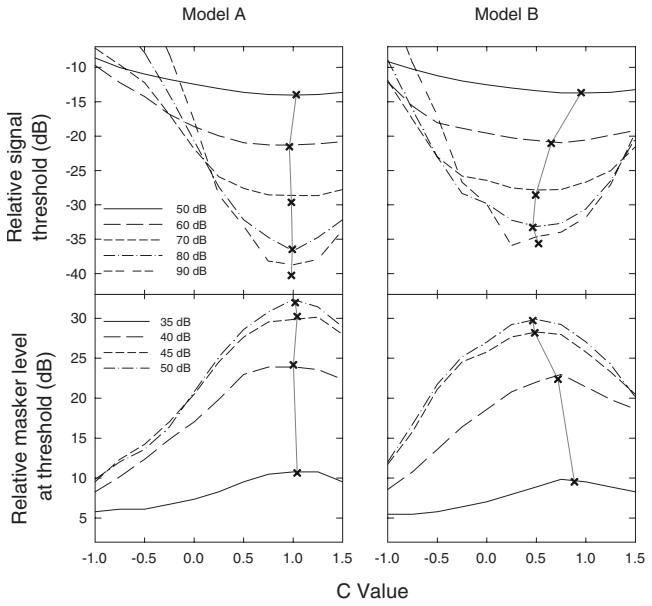


FIG. 7. The predictions from Model A (left) and Model B (right) for 2000-Hz conditions in Experiment 1 (upper panels) and Experiment 2 (lower panels). Different stimulus levels are indicated by different line styles. The predicted thresholds at each stimulus level were fit using a sinusoidal function to estimate the C value that led to the least effective masker. The minima (Experiment 1) and maxima (Experiment 2) of these fitted functions are marked with Xes and connected with gray lines.

vature. The model simulation will demonstrate that when a constant curvature across the auditory-filter passband is assumed, no shift in the psychophysical curvature estimates would be observed even for a level-varying magnitude response. In contrast, a shift of the behaviorally measured curvature similar to what has been observed in Experiments 1 and 2 could be achieved by having a level-invariant but frequency-variant phase curvature.

Following the approach of Oxenham and Dau (2001a), gammachirp filters with the magnitude responses described by Irino and Patterson (2001) and artificial phase characteristics are presented. The magnitude responses, phase responses, and group delays of these filters are shown in Fig. 6. The two models have the same magnitude response, which was adopted from the compressive gammachirp filters at various stimulus levels.[3] The center frequency of the auditory filter was fixed at 2 kHz. Model A has a constant auditory-filter curvature equivalent to $C=-1$ (corresponding to a dimensionless curvature of -16); hence it has a linear group delay function in the frequency domain. In contrast, Model B has zero curvature at low frequencies up to 1.75 kHz and a constant curvature equivalent to $C=-1$ at all other frequencies. The group delay is therefore a piece-wise linear function (see the bottom panel of Fig. 6). Note that 1.75 kHz is approximately 1 equivalent rectangular bandwidth (ERB) below the filter center frequency (2 kHz).

In order to provide model predictions of the experimental data, stimuli were the maskers used for testing at 2000 Hz. Maskers alone and maskers with a 2000-Hz signal were both passed through a single auditory filter (Model A or Model B)[4] in cascade with a model of hair cell and auditory nerve fibers (Meddis, 1986) forming internal representations of the stimuli. All model parameters (for either Model A or Model B), except the magnitude response, were fixed across all experimental conditions, with the different magnitude responses used at each level. An internal power ratio between the signal-plus-masker and the masker alone was calculated from the mean-square firing rates produced by the model.[5] The signal level at threshold was determined when the power ratio exceeded a certain criterion; this criterion was fixed across all conditions. Results from Models A and B are shown in the left and right panels of Fig. 7, respectively. The simulations of Experiments 1 and 2 are shown separately in the upper and the lower panels. In the simulation of Experiment 2, signal levels of 35-, 40-, 45-, and 50-dB SPL were used instead of the ones in the actual experiment. This was due to the limited dynamic range of the Meddis (1986) model, which could not provide reliable results at low stimulus levels. The curve-fitting procedure described in Experiments 1 and 2 was performed to estimate $C_{\min}$ from the predicted thresholds. For both experiments, predicted thresholds at each stimulus level were curve fitted using a sinusoidal function. The minima (Experiment 1) and maxima (Experiment 2) of these fitted functions are indicated in Fig. 7 with Xes.

Both models illustrate the same general trends that are present in the experimental data, but each model also reveals certain limitations. First, Models A and B accurately predict a substantial change in threshold with increasing C and all curves have pronounced minima (Experiment 1) or maxima (Experiment 2). Second, the two models show that a level-

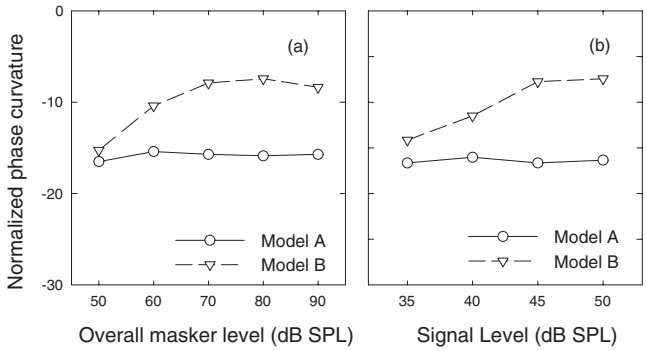


FIG. 8. The normalized magnitude of the phase curvature obtained for 2000-Hz conditions in (a) Experiment 1 and (b) Experiment 2. Results from Models A and B are shown as open circles and triangles, respectively.

dependent masker phase effect is present—as stimulus level increases the difference in masked threshold between $C=-1$ and $C_{\min}$ also increases. For Model A, $C_{\min}$ is always located at or close to $C=1.0$, the same value used to generate the curvature of the auditory-filter phase response. The clear correspondence between the two curvatures provides robust evidence that psychophysical experiments can be used to measure the auditory-filter phase curvature. However, Model A does not reveal a level-dependent shift in $C_{\min}$, as was observed in the data for both Experiments 1 and 2. In contrast, Model B leads to a level-dependent estimate of $C_{\min}$. The values of $C_{\min}$ range between 0.5 at the highest masker levels and 1.0 at the lowest masker levels. Notably, this factor of 2 change in $C_{\min}$ is quite similar to that observed in the experimental data. In this case, the model with a level-invariant and frequency-varying auditory-filter phase curvature predicts a level-dependent shift in $C_{\min}$. At the lowest stimulus level tested, the model prediction of $C_{\min}=1$ is consistent with the curvature used for the auditory-filter phase response at the CF. However, the model does not accurately predict the curvature of the phase response for the higher masker levels (e.g., 80- and 90-dB SPL). For these highest levels, the estimated phase curvature is likely due to greater weighting of low-frequency components at the output of the auditory filter. This greater weighting leads to an estimated curvature that is somewhere between zero (the phase curvature of the low-frequency tail of the auditory filter) and the curvature associated with $C=-1$ (the phase curvature in the auditory-filter passband). These modeling results demonstrate that the off-frequency phase response can influence the behaviorally estimated auditory-filter phase curvature, as suggested by Summers *et al.* (2003). Given this result, one must be cautious in interpreting psychophysical measured curvature as providing estimates of the true auditory-filter curvature, especially at higher stimulus levels.

To compare the resulting estimated curvatures of the two models more closely, the $C_{\min}$ values, converted to normalized curvature values, are plotted in Fig. 8. Figure 8 illustrates that Model A predicts a curvature that is more negative than Model B for all stimulus levels, and this predicted curvature is in good agreement with the auditory-filter curvature used for Model A (-16). The curvature magnitudes predicted by Model B also decrease with increasing level. At the lowest stimulus levels, the psychophysical curvature estimates are close to the value used in the auditory-filter model. When comparing the model curvature predictions with those of Experiments 1 and 2, it is apparent that Model B provides a better approximation to the experimental data of Experiment 1 (Fig. 3) and Experiment 2 (Fig. 5). Specifically, Model B leads to a fairly constant prediction of curvature for stimulus levels above 70-dB SPL (e.g. Experiment 1), which is comparable to the plateau of the estimated phase curvature observed in the data at 2000 Hz. The simplified phase curvature used here is likely to be similar but not necessarily identical to the phase curvature of the auditory filter. Further work characterizing the phase characteristics of the auditory-filter phase curvature may be necessary to yield better predictions of the data. Regardless, this model demonstrates that level-dependent psychophysical phase curvature estimates might reflect a level-independent auditory-filter phase curvature combined with level-dependent changes in the magnitude response with level.

These modeling results demonstrate that 1) if the auditory-filter curvature is level- and frequency-invariant (as in Model A), no shift in psychophysical curvature estimates would be observed with increasing stimulus level. 2) With the same modeling framework but a frequency-dependent auditory-filter curvature (as in Model B), shifts in the psychophysical curvature estimates similar to the ones observed in experimental data are apparent despite a level-invariant auditory-filter phase curvature. Simulations from Model B illustrate that the auditory-filter curvature at frequencies below 1 ERB from the auditory-filter center frequency could contribute to the psychophysical curvature estimates at high levels. Additional modeling shows that when the critical frequency dividing the two segments of the auditory-filter curvature (zero-curvature and constant-curvature segments) in Model B was set to 1.5 kHz (about 2 ERBs below the center frequency), the shifts in the psychophysical curvature with the increasing level could still be measured. It seems that the psychophysical curvature consists of integration of the auditory-filter curvature in these models.

V. SUMMARY

The present study provides estimates of the auditory-filter phase curvature at different stimulus levels using modified Schroeder-phase maskers. Two experiments, one using a fixed masker level and another using a fixed signal level, indicate that the estimated magnitude of the phase curvature decreases as stimulus level increases. This result demonstrates the existence of level dependence in the behaviorally measured phase curvature of the auditory filter, even though such results are not consistent with physiological measurements of the phase response at a single cochlear location. In an effort to resolve the contradictions between psychophysical and physiological results, a plausible mechanism underlying the psychophysical curvature is suggested following the "off-frequency phase influence" hypothesis by Summers *et al.* (2003), in which psychophysically estimated curvature reflects an interaction between the magnitude and the phase response of the auditory filter. Psychoacoustic modeling displays evidence of the viability of this mechanism. To develop

more accurate future behavioral techniques for estimating the auditory-filter phase curvature, the role of the magnitude response has to be carefully considered.

ACKNOWLEDGMENTS

We would like to thank Dr. Armin Kohlrausch and Dr. Andrew Oxenham for their constructive reviews and helpful suggestions. Dr. Toshio Irino provided the computer program for the implementation of the gammachirp filter. Data collection was assisted by Melissa Papesh and Susie Valentine, and Dr. Robert Withnell provided many fruitful discussions that have contributed to the development of this work.

[1]Rather than testing the same signal frequencies as Oxenham and Dau (2001b), 2-kHz signals were selected instead of 1-kHz or 250-Hz tones because of the following considerations: (1) Auditory-filter bandwidths are wider for higher frequencies, allowing a greater number of components to interact within a single auditory filter. These interactions will produce greater changes in threshold at low levels and will allow more accurate estimates of auditory-filter phase curvature. (2) At frequencies below 1 kHz, a significant portion of the masker power could be presented to cochlear locations where the logarithmic frequency mapping does not hold. This might confound the interpretation of the results, and such a confound will not be as pronounced for frequencies above 1 kHz. (3) The threshold minima for signal frequencies below 1 kHz are likely to exceed $C=1$ for most listeners for the chosen masker fundamental frequency and masker bandwidth in the experiment, therefore potentially making the estimation of the minimum difficult.

[2]The low-pass noise used in these spot checks had a cut-off frequency of 600 Hz and a spectrum level of 41-dB SPL for the 90-dB conditions and 11-dB SPL for the 60-dB conditions. These levels were chosen to be about 15 dB lower than the average spectrum level of the complex masker, following Oxenham and Dau (2001b).

[3]Model parameters were identical to those in the original paper (Irino and Patterson, 2001): $n=4$, $b_1=2.02$, $b_2=1.14$, $c_1=-3.70$, $c_2=0.979$, and $f_{\mathrm{rat}}=0.573+0.0101P_s$. The compressive gammachirp filter depends on the stimulus level through the parameter P_s, which is the probe tone level in the notched-noise experiment used for model fitting (Irino and Patterson, 2001). In the present study, P_s was set to be 20 dB below the masker level. This is a rough approximation equating the sound energy within the passband of the filter between our experiments and the notched noise paradigm.

[4]It is worth pointing out that Models A and B are functionally linear. Because the filter's magnitude and phase response are separately specified, the filters could be conveniently realized via linear FIR filters. The possibility of using linear filters to predict cochlear responses to Schroeder-phase stimuli has been studied by Summers *et al.* (2003) in detail. They measured a series of "indirect" impulse responses (de Boer and Nuttall, 1997) of the cochlea at various levels. Responses were predicted by convolving the acoustic Schroeder-phase stimuli with the indirect impulse responses. These modeled responses were compared with the experimentally measured responses. Results showed that realistic predictions could be achieved by this series of linear models, which gives justification in applying linear-filter models here.

[5]The durations of the stimuli used here were 300 ms. The first 100 ms and the last 50 ms of the model output was excluded from the calculation of the mean-square ratio of the firing rates to minimize the effect of the onset and offset transients.

Bacon, S. P., Lee, J., Peterson, D. N., and Rainey, D. (**1997**). "Masking by modulated and unmodulated noise: Effects of bandwidth, modulation rate, signal frequency, and masker level," J. Acoust. Soc. Am. **101**, 1600–1610.

Buus, S. (**1985**). "Release from masking caused by envelope fluctuations," J. Acoust. Soc. Am. **78**, 1958–1965.

Carlyon, R. P., and Datta, A. J. (**1997a**). "Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system," J. Acoust. Soc. Am. **101**, 3636–3647.

Carlyon, R. P., and Datta, A. J. (**1997b**). "Masking period patterns of Schroeder-phase complexes: Effects of level, number of components, and phase of flanking components," J. Acoust. Soc. Am. **101**, 3648–3657.

Carney, L. H., McDuffy, M. J., and Shekhter, I. (**1999**). "Frequency glides in the impulse responses of auditory-nerve fibers," J. Acoust. Soc. Am. **105**, 2384–2391.

de Boer, E., and Nuttall, A. L. (**1997**). ""The mechanical waveform of the basilar membrane. I. Frequency modulations ("glides") in impulse responses and cross-correlation functions," J. Acoust. Soc. Am. **101**, 3583–3592.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.

Irino, T., and Patterson, R. D. (**2001**). "A compressive gammachirp auditory filter for both physiological and psychophysical data," J. Acoust. Soc. Am. **109**, 2008–2022.

Kohlrausch, A., and Sander, A. (**1995**). "Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets," J. Acoust. Soc. Am. **97**, 1817–1829.

Lentz, J. J., and Leek, M. R. (**2001**). "Psychophysical estimates of cochlear phase response: Masking by harmonic complexes," J. Assoc. Res. Otolaryngol. **2**, 408–422.

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

Meddis, R. (**1986**). "Simulation of mechanical to neural transduction in the auditory receptor," J. Acoust. Soc. Am. **79**, 702–711.

Oxenham, A. J., and Dau, T. (**2001a**). "Reconciling frequency selectivity and phase effects in masking," J. Acoust. Soc. Am. **110**, 1525–1538.

Oxenham, A. J., and Dau, T. (**2001b**). "Towards a measure of auditory-filter phase response," J. Acoust. Soc. Am. **110**, 3169–3178.

Oxenham, A. J., and Dau, T. (**2004**). "Masker phase effects in normal-hearing and hearing-impaired listeners: Evidence for peripheral compression at low signal frequencies," J. Acoust. Soc. Am. **116**, 2248–2257.

Oxenham, A. J., and Ewert, S. D. (**2005**). "Estimates of auditory filter phase response at and below characteristic frequency," J. Acoust. Soc. Am. **117**, 1713–1716.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," J. Acoust. Soc. Am. **72**, 1788–1803.

Recio, A., and Rhode, W. S. (**2000**). "Basilar membrane responses to broadband stimuli," J. Acoust. Soc. Am. **108**, 2281–2298.

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.

Schroeder, M. (**1970**). "Synthesis of low peak factor signals and binary sequences with low autocorrelation," IEEE Trans. Inf. Theory **16**, 85–89.

Shera, C. A. (**2001**). "Frequency glides in click responses of the basilar membrane and auditory nerve: Their scaling behavior and origin in traveling-wave dispersion," J. Acoust. Soc. Am. **109**, 2023–2034.

Smith, B. K., Sieben, U. K., Kohlrausch, A., and Schroeder, M. R. (**1986**). "Phase effects in masking related to dispersion in the inner ear," J. Acoust. Soc. Am. **80**, 1631–1637.

Summers, V. (**2000**). "Effects of hearing impairment and presentation level on masking period patterns for Schroeder-phase harmonic complexes," J. Acoust. Soc. Am. **108**, 2307–2317.

Summers, V., de Boer, E., and Nuttall, A. L. (**2003**). "Basilar-membrane responses to multicomponent (Schroeder-phase) signals: Understanding intensity effects," J. Acoust. Soc. Am. **114**, 294–306.

Enhancing sensitivity to interaural time differences at high modulation rates by introducing temporal jitter

Matthew J. Goupell,[a] Bernhard Laback, and Piotr Majdak
Acoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14, A-1040 Vienna, Austria

(Received 26 June 2008; revised 18 June 2009; accepted 27 June 2009)

Sensitivity to interaural time differences (ITDs) in high-frequency bandpass-filtered periodic and aperiodic (jittered) pulse trains was tested at a nominal pulse rate of 600 pulses per second (pps). It was found that random binaurally-synchronized jitter of the pulse timing significantly increases ITD sensitivity. A second experiment studied the effects of rate and place. ITD sensitivity for jittered 1200-pps pulse trains was significantly higher than for periodic 600-pps pulse trains, and there was a relatively small effect of place. Furthermore, it could be concluded from this experiment that listeners were not solely benefiting from the longest interpulse intervals (IPIs) and the instances of reduced rate by adding jitter, because the two types of pulse trains had the same longest IPI. The effect of jitter was studied using a physiologically-based model of auditory nerve and brainstem (medial superior olive neurons). It was found that the random timing of the jittered pulses increased firing synchrony in the auditory periphery, which caused an improved rate-ITD tuning for the 600-pps pulse trains. These results suggest that a recovery from binaural adaptation induced by temporal jitter is possibly related to changes in the temporal firing pattern, not spectral changes.
 [DOI: 10.1121/1.3206584]

PACS number(s): 43.66.Pn, 43.66.Ba, 43.66.Qp, 43.66.Mk [RLF] Pages: 2511–2521

I. INTRODUCTION

The experiments described here were motivated by recent work on interaural time difference (ITD) sensitivity in cochlear-implant (CI) users and past work on the binaural adaptation phenomenon observed in ITD perception of normal-hearing (NH) listeners. Hafter and Dye (1983) presented evidence that ITD sensitivity decreases for high-frequency modulated stimuli, like bandpass-filtered pulse trains, if the modulation rate is too high. By systematically varying the number and rate of pulses in a train, they found that increasing the pulse rate decreases the usefulness of the binaural information after the onset. Later, Hafter and Buell (1990) showed that a recovery from binaural adaptation can be produced by inserting a change or "trigger" in the signal. They reported a recovery from adaptation when doubling or halving one or more intervals in a pulse train with a 2.5-ms interpulse interval (IPI). They also reported a recovery from binaural adaptation to a pulse train by adding short trigger signals such as diotic sinusoids or diotic, monotic, or uncorrelated noise bursts. Hafter and Buell (1990) concluded that the recovery effect from a trigger was most likely due to a temporary spectral change in the signal. The previously described studies were performed with a fixed number of pulses. In fact, even for a periodic stimulus with a fixed duration (such as pulse trains, sinusoidally amplitude modulated tones, or transposed tones), there is decreasing ITD sensitivity with increasing modulation rate (e.g., Bernstein and Trahiotis, 2002; Majdak and Laback, 2009).

Recent work with CIs readdressed the binaural adaptation phenomenon and introduced a new method to cause a recovery from binaural adaptation (Laback and Majdak, 2008). CIs use high-rate electric pulses to encode acoustic information. Several recent studies have shown that, similar to NH listeners, ITD sensitivity in CI listeners rapidly decreases with increasing pulse rate beyond a few hundred pulses per second (pps) (Majdak *et al.*, 2006; Laback *et al.*, 2007; van Hoesel, 2007). Laback and Majdak (2008) hypothesized that this pulse rate limitation is a form of binaural adaptation and showed that introducing binaurally-synchronized jitter (referred to as binaural jitter) can substantially increase ITD sensitivity at rates of 800–1515 pps. Because direct electric stimulation at one interaural electrode pair was used in that experiment, the jitter changed only the temporal properties of the stimuli, not the spectral. Therefore, they concluded that the recovery from binaural adaptation is caused by ongoing temporal changes in the signal.

In this study, we examined if a similar improvement in ITD sensitivity could be achieved in acoustic hearing by introducing binaural jitter into high-frequency bandpass-filtered pulse trains. In experiment 1, we tested the effect of binaural jitter for 600-pps pulse trains, a pulse rate at which listeners normally have difficulty in detecting waveform ITDs (Majdak and Laback, 2009). In experiment 2, we tested the hypothesis that the improvement in ITD sensitivity depends on only the longest IPIs of a jittered pulse train. We then modeled the response of the auditory periphery and brainstem to jittered pulse trains to observe the likely changes to the physiological firing patterns introduced by jitter in an attempt to understand the listeners' ITD sensitivity.

The effect of binaural jitter on ITD sensitivity has already been investigated in two earlier studies using sinusoids (Nordmark, 1976; Blauert, 1981). Nordmark (1976) reported surprisingly small just noticeable differences (JNDs) around

[a]Author to whom correspondence should be addressed. Electronic mail: goupell@wisc.edu

1.5 μs for a temporally-jittered 4-kHz carrier. Blauert (1981) replicated Nordmark's experiment and found JNDs that were two orders of magnitude larger (around 170 μs). If the latter measurement is correct, this means that ITD JNDs for jittered sinusoids are comparable to other high-frequency stimuli that have an amplitude modulation (AM) (Henning, 1974) or a frequency modulation (FM) (Henning, 1980). This might be expected if a jittered sinusoid is viewed as a FM with a random modulation frequency. Note, however, the fundamental difference between the study by Nordmark (1976) and Blauert (1981) using jittered sinusoids, and both Laback and Majdak's (2008) study and the present study using pulse trains. The purpose of using jittered sinusoids in earlier studies was to present usable ITD information at high center frequencies. The purpose of this study is to investigate the effect of jitter on ITD rate limitations for pulsatile stimuli, which are commonly associated with CI processing strategies, and to more deeply understand how temporal jitter affects ITD sensitivity.

II. EXPERIMENT 1

A. Listeners and equipment

Six listeners participated in this experiment. All listeners were between 24 and 37 years old and had normal hearing according to standard audiometric tests. Two listeners were authors of this study (NH2 and NH10). All six listeners were experienced with virtual sound localization. Three listeners (NH2, NH8, and NH10) had extensive experience in lateralizing pulse trains. From preliminary tests and training, we determined that listeners NH2, NH8, and NH10 could lateralize pulse trains with relatively small amounts of jitter compared to the other listeners. Therefore, we divided the listeners into a high-sensitivity group (NH2, NH8, and NH10) and a low-sensitivity group (NH12, NH14, and NH15). Also, NH8 was markedly more sensitive to ITD than the other five listeners. Thus, he was given smaller ITD values to lateralize to avoid ceiling effects.

A personal computer system was used to control the experiment. The stimuli were output via a 24-bit stereo A/D-D/A converter (ADDA 2402, Digital Audio Denmark) using a sampling rate of 96 kHz/channel. The analog signals were sent through a headphone amplifier (HB6, TDT) and an attenuator (PA4, TDT). The signals were presented to the subjects via headphones (HDA200, Sennheiser). Calibration of the headphone signals was performed using a sound level meter (2260, Brüel & Kjær) connected to an artificial ear (4153, Brüel & Kjær).

B. Stimuli

The stimuli were 500-ms pulse trains composed of 10.4-μs monophasic pulses, corresponding to one sampling interval at a sampling rate of 96 kHz. The pulse rate was 600 pps, which has an IPI of 1667 μs. A recent study by Majdak and Laback (2009) showed that ITD sensitivity degrades to chance around 500 pps for most NH listeners, which is generally consistent with studies that use other types of modulated stimuli (e.g., Bernstein and Trahiotis, 2002). Thus, stimuli with a 600-pps pulse rate have the property that there is substantial room for improvement in ITD sensitivity.

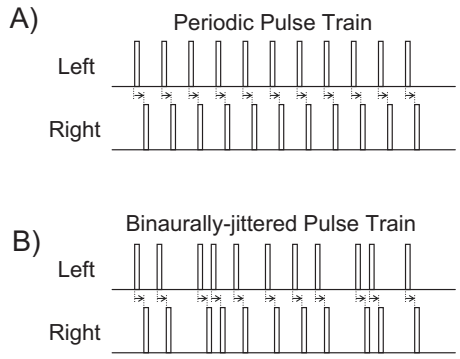


FIG. 1. Examples of a periodic pulse train (A) and a binaurally-jittered pulse train (B). The arrows indicate that the ITD of the two signals is always constant between pulses, independent of the IPI of the pulses.

A waveform ITD was introduced by delaying the temporal position of the pulses at one ear relative to the other ear. The ITD values were 100, 200, 400, and 600 μs for all but one listener. The other listener, who was unusually sensitive to ITDs, was tested with ITD values of 20, 50, 100, and 150 μs. To minimize the detection of ITD in the onset and offset of the stimulus, 150-ms linear ramping was applied to the pulse trains. The full-on duration of the stimuli was 200 ms. The −3-dB duration of the stimuli was 288 ms.

Jitter in the timing of the pulses was applied to the stimuli. Periodic pulse trains [Fig. 1(A)] had a constant IPI, whereas the jittered pulse trains [Fig. 1(B)] had randomly-varied IPIs. The nominal IPI corresponded to the average IPI over the stimulus duration. To preserve the ITD information in the pulse timing, the jitter was synchronized between the two ears (indicated by the constant length of the arrows in Fig. 1). The jitter followed a rectangular distribution, where the parameter k defines the width of the distribution relative to the nominal IPI. The parameter k ranges from 0 (periodic, no jitter) to 1 (maximum jitter). A jittered pulse train was "constructed" pulse by pulse. For each pulse added, the IPI was varied within the range of IPI$\cdot(1\pm k)$. Thus, for $k=1$, the largest possible IPI was twice the nominal IPI and the smallest possible IPI was zero. For the three high-sensitivity listeners, the jitter values were $k=0$ (periodic condition, no jitter), 1/128, 1/32, 1/8, and 1/3. For the three low-sensitivity listeners, the jitter values were $k=0$, 1/8, 1/3, 1/2, and 3/4.[1] Each trial used a new random jitter manifestation.

The pulse trains were passed through a digital sixth-order bandpass Butterworth filter. The spectral center frequency of the band was 4.6 kHz. The spectral bandwidth was 1.5 kHz. The A-weighted sound pressure level of the stimuli was 72 dB (re: 20 μPa). In a control condition, Gaussian white noises were used as stimuli, filtered by the same sixth-order Butterworth bandpass filter that was used for the jittered pulse trains.

Binaurally-uncorrelated, low-pass filtered, white noise was used to mask low-frequency components that might contain useful binaural cues. The corner frequency was

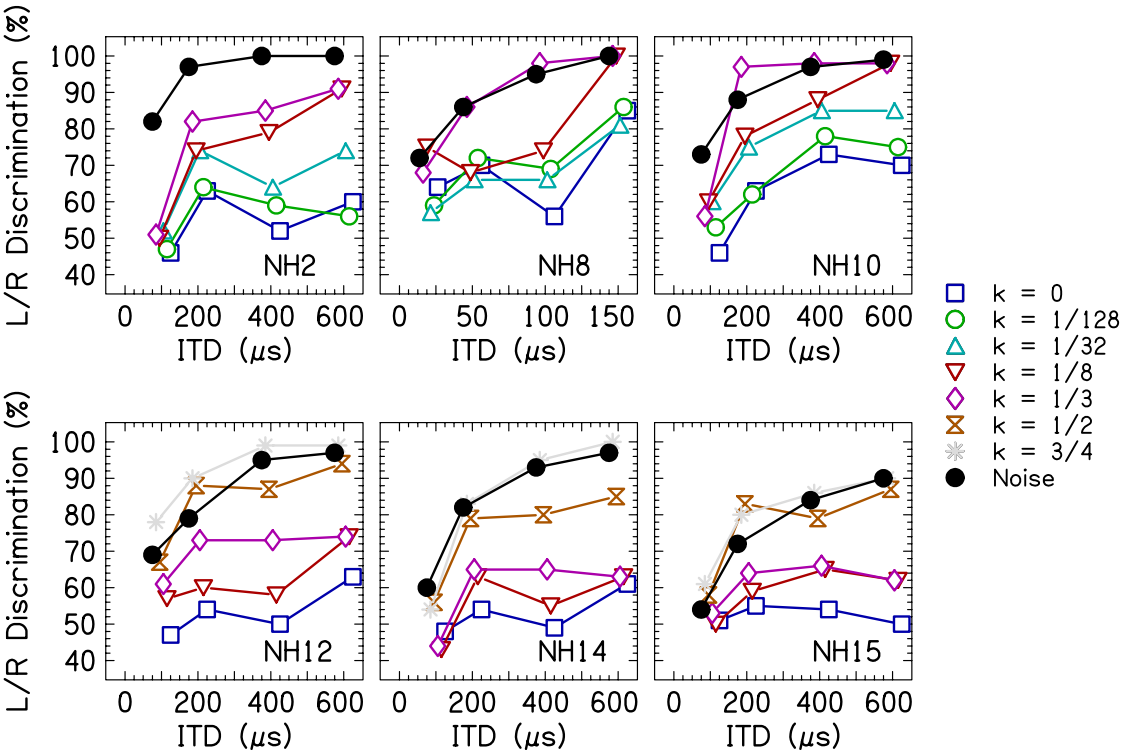


FIG. 2. (Color online) Results from experiment 1, the percentage of correct left-right discriminations for pulse trains with various jitter values and for noises. The high-sensitivity listeners are plotted in the top row, the low-sensitivity listeners in the bottom. Listener NH8 has smaller values of ITD than the other five listeners.

3500 Hz, with a 24-dB/oct roll-off, and the *A*-weighted sound pressure level of the noise was 61 dB. The sound pressure spectrum level at 2 kHz was 35.8 dB (re: 20 μPa in a 1-Hz band).

C. Procedure

A two-interval, two-alternative forced-choice procedure was used in a lateralization discrimination test. The first interval contained a reference stimulus with zero ITD and zero k evoking a centralized auditory image. The second interval contained the target stimulus with non-zero ITD and one of the five values of k. The interstimulus interval was 400 ms. The listeners indicated whether the second stimulus was perceived to the left or to the right of the first stimulus by pressing a button. Visual response feedback was provided after each trial. The listeners controlled when the next trial began.

For the pulse trains, a block contained 2000 trials consisting of 100 presentations of four ITD and five k values in a randomized order. For the noises, a block contained 400 trials consisting of 100 presentations of four different values of ITD. The 100 repetitions per condition were presented in a balanced format with 50 targets on the left and 50 targets on the right. The chance rate was 50%. Listeners took a break every 200–250 trials, which was approximately every 15 min. The order of the two blocks was balanced over listeners.

Listeners were trained before the main test started. The training began with stimuli that had $k=1/3$ and ITD=600 μs. Values of k and/or ITD were decreased as listeners' performance improved. Training continued until performance saturated. Listeners were separated into high-sensitivity and low-sensitivity groups after the training. If a listener could left/right discriminate $k=1/3$, ITD=200-, 400-, and 600-μs pulse trains more than 80% of the time, they were placed within the high-sensitivity group. The training period lasted between 4 and 8 h depending on the listener.

D. Results

Figure 2 shows the results of the experiment. The high-sensitivity listeners (NH2, NH8, and NH10) are plotted in the top row. The low-sensitivity listeners (NH12, NH14, and NH15) are plotted in the bottom row. Several of the psychometric functions asymptote well below 100% correct. Several of the listeners show a decrease in percent correct (P_c) for the periodic ($k=0$) 400-μs condition where the ITD is ambiguous for a 600-pps pulse train. The data show that adding jitter to pulse trains increases ITD discrimination performance and eliminates the decreases in P_c due to the ITD ambiguity. The performance for the jittered pulse trains seems to be approximately limited by the performance for the bandpass-filtered noise stimuli.

A two-way repeated-measures analysis of variance (RM ANOVA) (factors: ITD and k) was performed. The values of P_c were transformed using the rationalized arcsine transform proposed by Studebaker (1985) to not violate the homogeneity of variance assumption required for an ANOVA. The RM ANOVA showed that the main effects were highly significant ($p<0.0001$ for both), but the interaction was not statistically significant ($p=0.47$). Tukey HSD *post-hoc* tests were performed separately for the two listener groups to determine

TABLE I. JNDs calculated for experiment 1. JNDs that were not determinable are labeled "ND." JNDs that were not measured for a given listener are labeled "—."

	High-sensitivity			Low-sensitivity		
k	NH2	NH8	NH10	NH12	NH14	NH15
0	ND	83.9	442.6	ND	ND	ND
1/128	ND	65.0	329.2	—	—	—
1/32	396.2	89.9	152.9	—	—	—
1/8	253.1	45.1	138.0	ND	ND	ND
1/3	165.6	27.2	131.8	214.4	ND	ND
1/2	—	—	—	78.1	137.6	167.0
3/4	—	—	—	68.9	128.7	155.5
Noise	60.3	22.3	81.6	112.6	201.8	132.9

the value of k that shows a significant increase from the periodic condition. For the high-sensitivity listeners, $k=0$ did not differ from $k=1/128$ ($p=0.94$) and $k=1/32$ ($p=0.22$); $k=0$ significantly differed from $k=1/8$ ($p=0.0004$) and $k=1/3$ ($p<0.0001$). For the low-sensitivity listeners, $k=0$ did not differ from $k=1/8$ ($p=0.30$); $k=0$ significantly differed from $k=1/3$ ($p=0.001$), $k=1/2$ ($p<0.0001$), and $k=3/4$ ($p<0.0001$).

To more easily compare our results to those of previous studies, JNDs were estimated for each listener.[2] The threshold criterion was set to 70% and JNDs are reported in Table I. Some JNDs could not be computed because there were no P_c values above 70%.

E. Discussion

The data in Fig. 2 show that introducing jitter to the pulse timing of an acoustic pulse train can substantially improve ITD discrimination performance. Depending on the sensitivity of the listener, the amount of jitter that increased ITD sensitivity was different. The high-sensitivity listeners showed significant improvements for jitter values as small as 1/8. The low-sensitivity listeners showed significant improvements for jitter values as small as 1/3.

Listeners showed no or low sensitivity to ITD in the periodic 600-pps pulse trains, which was expected based on pilot tests. In many cases for low values of k, JNDs could not be determined (ND in Table I), consistent with previous studies of ITD sensitivity at this rate (Majdak and Laback, 2009). In contrast to our results, studies using comparable-rate pulse trains reported determinable JNDs (Hafter and Dye, 1983; Dye and Hafter, 1984). This difference is most likely due to the fact that we used long (150-ms) temporal ramping and thus avoided onset cues, which are known to be important at such high rates (e.g., Saberi and Perrott, 1995; Laback *et al.*, 2007).

Binaurally-jittered pulse trains have not been tested before in acoustic hearing. Hafter and Buell (1990) tested the effect of inserting one or three gaps of 5 or 7.5 ms in a regular pulse train with a standard IPI of 2.5 ms and observed improvements of ITD JNDs as large as a factor of 2. Even though this modification has some similarities with binaural jitter, a direct comparison to our results is hindered by the several differences in the stimuli, including the pulse rate, the signal duration, and the manner of IPI modification. Laback and Majdak (2008) reported the effect of binaural jitter in electric pulse trains presented to CI listeners. They observed large improvements in ITD sensitivity similar to the improvements of the NH listeners in the current study. Again, a quantitative comparison between the two studies is hindered by differences in the stimuli, most importantly the difference between acoustic and electric hearing, but also the different pulse rates and the fact that the electric stimuli intentionally included a slowly-varying envelope modulation.

Nordmark (1976) and Blauert (1981) measured ITD sensitivity to jittered sinusoids, which can produce random AM at the output of some auditory filters as a result of FM-to-AM conversion. Thus, the jittered sinusoids may have similar temporal characteristics as our jittered pulse trains. Comparison of our JNDs to those for the previous two studies agrees with Blauert's measurement, who found an average JND of 173 μs for 5% jitter. For our experiment, the average JND was 213 μs for the high-sensitivity listeners for $k=1/32=3\%$ jitter. The JNDs were not determinable for the low-sensitivity listeners for this jitter value. Our measurements may be slightly larger compared to Blauert because we used low-frequency masking noise, which could have increased JNDs (Bernstein and Trahiotis, 2004).

Blauert (1981) measured an average JND of 35 μs for 1-octave noise centered at 4 kHz. Our average JND was 102 μs for a 1/2-octave noise. Assuming increasing sensitivity with increasing spectral bandwidth (Bernstein and Trahiotis, 1994), our JNDs are expected to be larger than Blauert's JNDs. As mentioned before, we included a low-frequency masking noise, which could have further increased JNDs.

The results show that ITD sensitivity increases as the amount of jitter increases. This gain appears to be limited to the performance achieved with the bandpass-filtered noise stimuli. A discussion of the similarities between noise and jittered pulse trains is provided in the general discussion.

III. EXPERIMENT 2: HIGHER RATE AND PLACE

By introducing jitter, portions of the pulse trains have a relatively long instantaneous IPI, which decreases the instantaneous rate. At high center frequencies, low-rate modulated stimuli are easier to lateralize than unmodulated stimuli

(Henning, 1974), or high-rate modulated stimuli (Bernstein and Trahiotis, 2002). It could be that the increase in ITD sensitivity with jitter was due to the listeners more effectively utilizing the long IPIs in the pulse train compared to the short IPIs. This hypothesis has two forms. The first form is that listeners utilized the IPIs longer than some critical absolute value. The second form is that listeners utilized the IPIs relatively longer than the surrounding IPIs. In this experiment, we tested the first form of this hypothesis. To do this, we used periodic 600-pps pulse trains and jittered 1200-pps pulse trains ($k=1$). If the performance at 1200 pps with jitter exceeds the performance at 600 pps without jitter, then the absolute length of the IPI cannot be the sole signal property that causes the increased ITD sensitivity with increasing jitter. This is because the maximum IPI for a 1200-pps pulse train with $k=1$ is precisely the IPI for a 600-pps pulse train without jitter.

To keep the number of resolved harmonics constant and the spectral bandwidth approximately constant in terms of critical bands [an equivalent rectangular bandwidth of approximately 2.3 at both center frequencies (Moore and Glasberg, 1983)], the spectral center frequency and bandwidth of the 1200-pps stimulus were increased by a factor of 2 relative to the 4.6-kHz pulse train. As control conditions, a 600-pps pulse train was tested at a 9.2-kHz center frequency and a 1200-pps pulse train was tested at 4.6 kHz. This allowed us to further study the effects of the rate and place parameters.

A. Methods

This experiment used the same methods as experiment 1 and tested three conditions. The first condition used stimuli that had a spectral center frequency of 4.6 kHz and a bandwidth of 1.5 kHz, like those of experiment 1, but with a pulse rate of 1200 pps. The jitter value was either $k=0$ or 1. The second and third conditions used stimuli that had a center frequency of 9.2 kHz and a bandwidth of 3 kHz. The second condition had 600-pps stimuli with $k=0$ or $1/3$ for the high-sensitivity listeners or $k=0$ or $3/4$ for the low-sensitivity listeners. These values of k matched the largest values of k for the 600-pps data tested in experiment 1 for particular listener groups. The third condition had 1200-pps stimuli with $k=0$ or 1 for all the listeners. The A-weighted sound pressure level was 72 dB for all of the stimuli. The masking noise was the same as used in experiment 1. The same six listeners participated in this experiment. For listener NH8, ITD values of 50 and 100 μs were tested. For the other listeners, the ITD values of 200 and 400 μs were tested.

B. Results

Figure 3 shows the results for experiment 2. The 4.6-kHz, 600-pps data are repeated from experiment 1. From the figure, it can be easily seen that for any specific condition the performance for the jittered pulse trains was always greater than for the periodic pulse trains. For the periodic conditions, there appear to be substantial floor effects as most of the P_c values are near 50%. For some jittered conditions and listeners, there appear to be some ceiling effects.

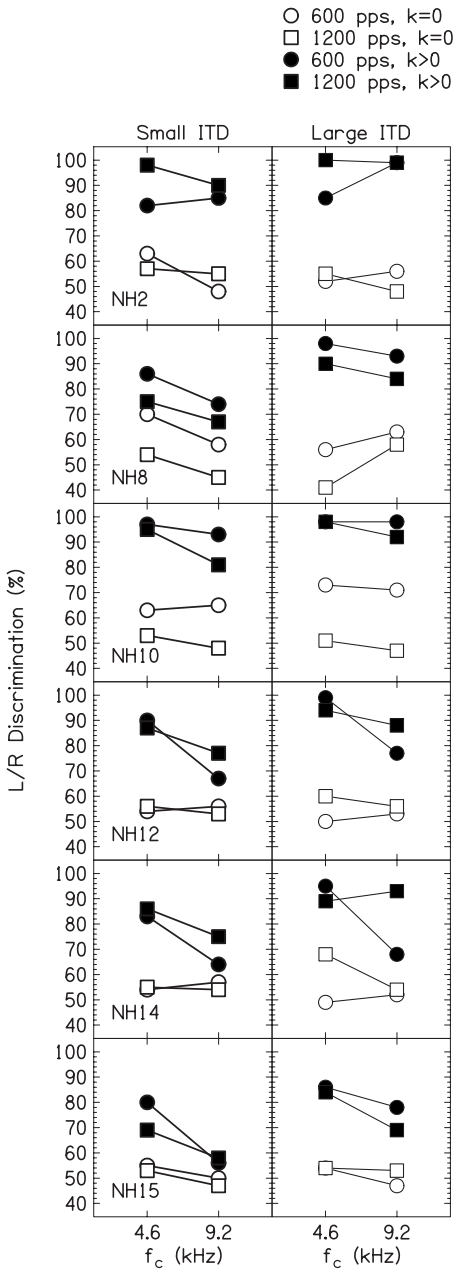


FIG. 3. Results from experiment 2. The high-performance listeners are plotted in the top three rows, the low-performance in the bottom three. Listener NH8 had smaller values of ITD (small ITD=50 μs and large ITD =100 μs) than the other five listeners (small ITD=200 μs and large ITD =400 μs). The high-sensitivity listeners were presented jittered 600-pps pulse trains with $k=1/3$. The low-sensitivity listeners were presented jittered 600-pps pulse trains with $k=3/4$. All six listeners were presented jittered 1200-pps pulse trains with $k=1$. The results for the 4.6-kHz 600-pps pulse trains are replotted from experiment 1.

To determine the effects of place and rate, specific cases were compared with a RM ANOVA. The values of $k=1/3$, $3/4$, or 1 were considered as a single condition with jitter for the statistical analysis. This is reasonable because the high-performance listeners in experiment 1 seemed to show a saturation of performance (approximately the performance for the bandpass noise) at $k=1/3$ and the low-performance listeners at $k=3/4$. It is assumed that a value of $k=1$ would only marginally improve the performance. First, one of the most important comparisons for this experiment is between

jittered 1200-pps pulse trains at 9.2 kHz and periodic 600-pps pulse trains at 4.6 kHz. There was a significant difference between these two conditions ($p<0.0001$). This indicates that the absolute length of the IPIs due to jittering pulse trains does not cause the observed improvements in ITD sensitivity. However, the comparison between these two conditions might be confounded by the effect of the different places and bandwidths.

The effect of place was tested using a RM ANOVA including only conditions with jitter to avoid floor effects. There was a significant decrease in performance with increasing place ($p=0.001$). Thus, even though there was an effect of place, it does not confound the conclusion of a larger performance for the jittered 9.2-kHz pulse trains compared to the periodic 4.6-kHz pulse trains. Rather, the effect of place reduces the difference.

It is possible that the increased sensitivity for the jittered 1200-pps pulse trains at 9.2 kHz compared to the periodic 600-pps pulse trains at 4.6 kHz is due to the increased spectral bandwidth used at 9.2 kHz. Therefore, we compared jittered 1200-pps pulse trains to periodic 600-pps pulse trains for a fixed place of 4.6 kHz. There was a significant difference between these two conditions ($p<0.0001$). Since the jittered 1200-pps pulse trains showed a higher performance than the periodic 600-pps pulse trains for a constant spectral bandwidth, it stands to reason that it was not the increased bandwidth that increased the performance when the place was changed.

C. Discussion

This experiment tested the hypothesis that jitter increases ITD sensitivity because it increases some IPIs beyond some absolute duration. Long IPIs may provide a benefit to listeners due to the refractoriness of some auditory neurons. This hypothesis can be rejected because it was shown that it was much easier to lateralize jittered 1200-pps pulse trains than periodic 600-pps pulse trains while systematically varying the rate and place parameters. Varying both parameters was necessary because by changing the rate, the number of resolved harmonics in the stimulus changed. The comparisons showed that place and rate had a comparatively small effect on ITD sensitivity compared to the jittering of pulse timing.

IV. MODELING OF NEURAL RESPONSE

Physiological measurements of responses of ventral cochlear-nucleus (VCN) chopper cells to maximum length sequence pulse trains (essentially jittered pulse trains) have been performed by Burkard and Palmer (1997). Their measurements showed that jitter increases the probability of a VCN neuron firing at certain time instances. Although spherical bushy VCN cells, not chopper VCN cells, project to the medial superior olive (MSO) (Smith *et al.*, 1993), insight may be gained from Burkard and Palmer's (1997) measurements. We hypothesized that the responses of the auditory nerve (AN) fibers, the input of the VCN, would become more synchronous after the introduction of jitter. We also hypothesized that increased synchrony will cause a sharpening of rate-ITD tuning. We assumed that sharpening of rate-ITD tuning is related to improvements in ITD sensitivity. Therefore, we modeled AN and MSO responses to binaurally-jittered pulse trains.

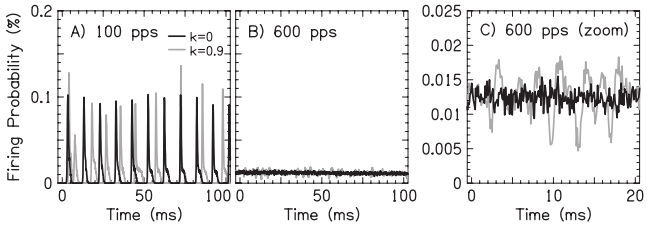


FIG. 4. Example PSTHs for 100-pps [panel (A)] and 600-pps pulse trains [panels (B) and (C)]. Stimuli are either periodic ($k=0$) or jittered ($k=0.9$). Only one manifestation is shown for each type of stimulus. The examples shown in panels (A) and (B) are from a 100-ms section from the steady state portions of the different types of stimuli. Panel (C) shows a 20-ms section for the 600-pps pulse train with a different vertical scale.

A. AN model

The model of the auditory periphery that was used was developed by Meddis (2006). We will briefly describe the model. A physical acoustic stimulus was filtered by a human outer and middle ear model based on Huber *et al.* (2001). The filtering of a human basilar membrane was modeled with a dual-resonance non-linear filter with parameters based on Tables II and III in Lopez-Poveda and Meddis (2001). The inner-hair cell (IHC) cilia, IHC presynapse, and AN synapse parameters were from Tables II, III, and IV of Meddis (2006), respectively. Only high-spontaneous rate fibers were modeled. The refractory time for the AN fibers was 0.75 ms. The model sampling rate was 10 kHz. The input stimuli had the same parameters (level, duration, rise-fall time, etc.) as the stimuli used in the experiments.

Figure 4 shows sample post-stimulus time histograms (PSTHs) for periodic and jittered ($k=0.9$) 100-pps and 600-pps pulse trains with a 4.6-kHz center frequency. The auditory filter was centered at 4.6 kHz. The 100-pps pulse trains show synchronous responses to both the periodic and jittered conditions, and there is little noticeable difference in the PSTHs with the exception for the expected aperiodic timing of peaks for the jittered pulse train. In contrast, for the periodic 600-pps pulse trains, the synchrony is not evident. Additionally, the jittered 600-pps pulse train shows noticeably higher peaks in the PSTH compared to the periodic 600-pps pulse train, which can easily be seen in Fig. 4(C).

We measured the firing rate and synchrony of the AN fibers' responses. Each measurement was made over 50 unique PSTHs. All calculations were made over the entire 500-ms stimulus duration. Figure 5 shows the average firing rate and the correlation index (CIn) (described below) for the average of five pulse train manifestations for five pulse rates (100, 300, 600, 900, and 1200 pps) as a function of k (0, 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, and 1). The responses of two filters with best frequencies (BFs) of 4.6 and 9.2 kHz were modeled. The AN firing rates for the 100-pps pulse trains are around 50–70 spikes/s. Theoretically, the firing rate should be near 100 spikes/s. As stated before, because of the long temporal onset and offset ramps, the −3-dB duration of the

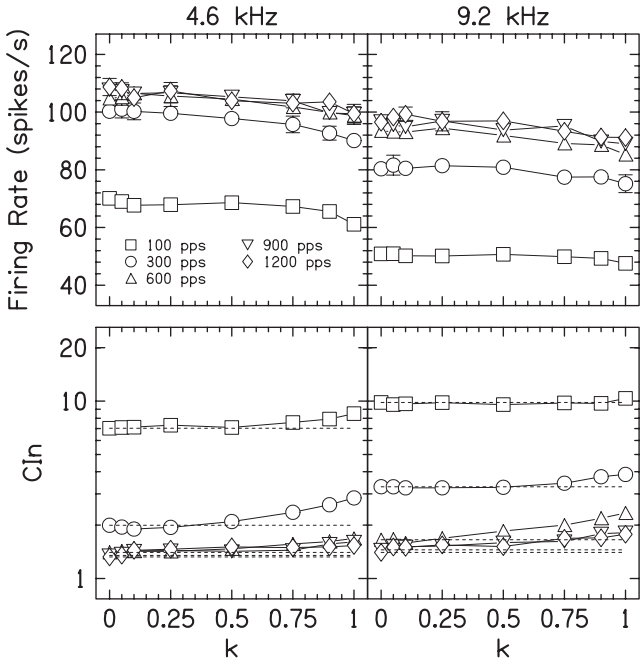


FIG. 5. The average AN firing rates and the CIn for five pulse rates as a function of k. The left panels are for the 4.6-kHz pulse trains. The right panels are for the 9.2-kHz pulse trains. Each point shows the average AN firing rate or CIn over the five manifestations of jitter. The error bars show ± 1 standard deviation over the five manifestations. The dotted lines show the value of the CIn for $k=0$ to more easily identify changes for increasing k.

stimuli was only 288 ms. If the temporal ramping was omitted from the input stimuli, the firing rates would be higher. The other pulse rates have a higher firing rate, limited by the refractory time of the AN fibers. For both center frequencies and all pulse rates, the AN firing rate decreases slightly with jitter. For the higher pulse rates, a small decrease in firing rate may be expected with jitter because, after short IPIs, pulses will be missed because of the refractory effects. This will not necessarily be compensated by the longer IPIs because it will depend on the lengths of the surrounding IPIs.

We quantitatively measured the change in synchrony when jitter was added to a pulse train. Since jittered pulse trains are aperiodic, a common metric like the synchronization index is not appropriate. Instead, we used a metric to allow for aperiodic stimuli, called the CIn (Joris *et al.*, 2006). The CIn is based on the counting of neural response spike coincidences from multiple presentations of the same stimulus. Mathematically, the CIn is

$$\mathrm{CIn} = \frac{2N_c}{M(M-1)r\omega T},$$

where N_c is the number of individual neuron firing coincidences, r is the average firing rate, M is the number of presentations, ω is the coincidence window duration, and T is the duration of the stimulus. The factor of 2 is necessary because we used the number of unordered pairs for our calculation, not the number of ordered pairs as in Joris *et al.* (2006). For our modeling, we used $\omega=100$ μs and $M=50$ presentations. The CIn has a value of 1 for an uncorrelated response, a value greater than 1 for a correlated response, and a value of 0 for an anticorrelated response.

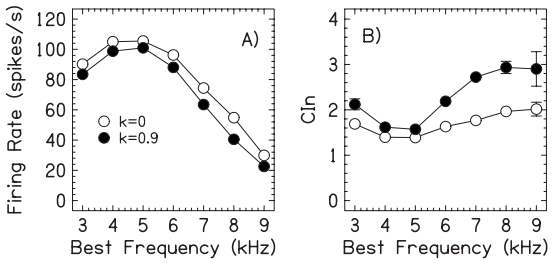


FIG. 6. AN firing rate and CIn as a function of auditory filter BF. The stimuli were periodic ($k=0$) and jittered ($k=0.9$) 600-pps pulse trains with a 4.6-kHz center frequency. Each point shows the average AN firing rate or CIn over the five manifestations of jitter. The error bars show ± 1 standard deviation over the five manifestations. Several points have error bars smaller than the size of the data point.

The bottom row of Fig. 5 shows the CIn. The dotted lines show the CIn for $k=0$. For an increase in jitter, both center frequencies and all pulse rates show an increase in CIn, hence more synchronous firing. The 100- and 300-pps pulse trains show increases for k greater than or equal to 0.5. In contrast, the higher pulse rates show an increase in CIn for values of k as small as 0.05. For a condition that showed a significant increase in ITD sensitivity in experiment 1 (high-sensitivity listeners), namely, the 600-pps pulse trains for $k=1/8=0.125$, there is an increase in firing synchrony.

Blauert (1981) postulated that the increase in ITD sensitivity to jittered sine tones was due to FM-to-AM conversion of the signal, which may happen due to the steep slopes of the auditory bandpass filters. He postulated this especially for off-frequency filters toward higher frequencies. To investigate the use of off-frequency cues for ITD sensitivity, we modeled the response of auditory filters with BFs from 3 to 9 kHz. We used periodic and jittered ($k=0.9$) 600-pps pulse trains, all with a 4.6-kHz spectral center frequency. Like before, we averaged our results over five different jitter manifestations. Figure 6 shows the results of varying the BF of the auditory filter. The jittered pulse trains always have a slightly smaller firing rate than the periodic pulse trains for all auditory filters modeled, although this difference is approximately constant for all BFs. The jittered pulse trains always have a larger CIn than the periodic pulse trains for all auditory filters modeled. The largest differences were for auditory filters away from the center frequency of the stimulus, in line with Blauert's (1981) notion that the AM in off-frequency filters could be important for detecting ITDs. Another explanation for the larger CIn (hence improved synchrony) with increasing auditory filter BF would be that the basilar membrane impulse response becomes shorter with increasing center frequency because the auditory filter bandwidth increases. The importance of the length of the basilar membrane impulse response is also supported by the results for periodic pulse trains in Fig. 5; the CIn for the 9.2-kHz band is consistently higher than the CIn for the 4.6-kHz band.

B. MSO coincidence model

Because we observed an increase in AN firing synchrony with an increase in jitter, we wondered if such an

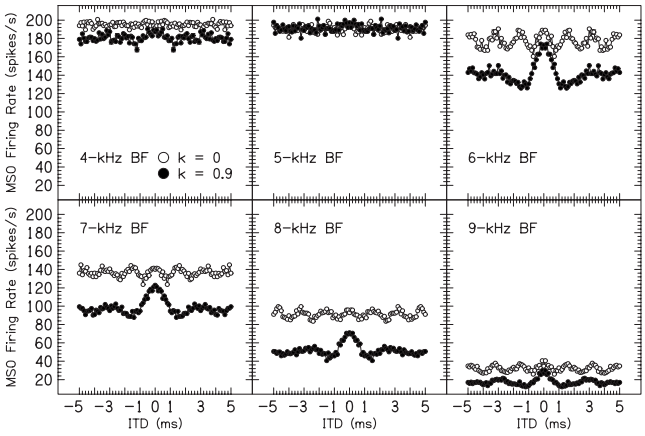


FIG. 7. MSO rate-ITD tuning curves for auditory filters with differing BF. The stimuli were periodic ($k=0$) and jittered ($k=0.9$) 600-pps pulse trains with a 4.6-kHz center frequency. Each point shows the average firing rate or CIn over the five manifestations of jitter. Error bars representing ± 1 standard deviation are all smaller than the data points.

increase could be utilized by the MSO to improve ITD perception. We simply used the response of the AN as the MSO input because it is presently unknown exactly how primary-like spherical bushy VCN cells alter the AN firing pattern. The modeled MSO neuron was a simple excitatory-excitatory coincidence counter. Thirty excitatory synapses (15 per side) provided the input to the cell. The cell fired if there was a coincident firing from the left and right inputs within a 100-μs window. One-hundred unique PSTHs were made and 30 (15 for each side) were randomly selected without replacement as the input to the MSO cell.[3] After a coincidence, the cell went into a refractory state where no firing occurred for 1 ms (Scott *et al.*, 2005; Scott *et al.*, 2007).[4] Each MSO measurement was repeated 100 times using different random sets of 30 PSTHs, chosen from the same pool of 100 PSTHs.

To show how an increased firing synchrony could translate to increased ITD sensitivity, we calculated MSO firing rates for binaural AN inputs with a range of ITDs. To support the psychophysical data, we expect sharper rate-ITD tuning for jittered pulse trains. The model MSO neuron had a best ITD of 0 μs. Figure 7 shows the rate-ITD curves for responses of auditory filters with BFs between 4 and 9 kHz. The input stimuli were a periodic and a jittered ($k=0.9$) 600-pps pulse train with a 4.6-kHz center frequency. Little difference in the tuning could be seen between the shapes of the curves for the periodic and jittered pulse trains at 4 and 5 kHz. At higher BFs, the periodic nature of the rate-ITD tuning curves is apparent for the periodic pulse trains. This periodic nature is not seen for the jittered pulse trains. As expected, because the CIn is larger for the jittered pulse trains, there was sharper rate-ITD tuning. The MSO firing rate decreases for all ITDs for best auditory filter frequencies of 6 kHz and higher. This is due to the decreasing firing rate in the AN with increasing BF, seen in Fig. 6(A).

V. GENERAL DISCUSSION

These experiments were inspired by previous studies on the binaural adaptation phenomenon (Hafter and Dye, 1983; Hafter and Buell, 1990). The results show that introducing binaurally-synchronized jitter into the timing of high-frequency filtered pulse trains considerably improves ITD sensitivity of NH listeners, consistent with the hypothesis that a change in the ongoing signal causes a recovery from binaural adaptation.

However, while Hafter and Buell (1990) and Hafter (1997) argued that the recovery effect is mediated by a discernible short-term change to the spectrum, the study by Laback and Majdak (2008) on the effect of binaural jitter on ITD sensitivity in CI listeners suggests that temporal changes alone can cause a recovery. Direct electrical stimulation at a single interaural electrode pair allowed the introduction of jitter to the pulse timing without concomitant spectral changes. The improvements in ITD sensitivity by binaural jitter in electrical hearing were similar to those observed in the present study with acoustic hearing. Of course, it is possible that different mechanisms are responsible for the improvements in electric and acoustic hearing. Hence, we cannot entirely dismiss the possibility that spectral changes also contributed to the recovery effect in acoustic hearing. It is worth noting that jitter, created by randomly modulating the rate of pulses, will cause an additional AM signal due to FM-to-AM conversion from auditory filtering in NH listeners. In contrast, in CI listeners, the auditory filters are bypassed. However, additional AM is probably created in both NH and CI listeners in the auditory system via synaptic transmission properties and neural membrane time constants.

In experiment 2, we hypothesized that the improvement in ITD sensitivity was due to only the introduction of long IPIs. Long IPIs could reduce the instantaneous modulation rate below the lowpass cutoff of a modulation filter or allow temporary recovery from refractoriness in auditory neurons. The data showed higher performance for jittered 1200-pps pulse trains compared to periodic 600-pps pulse trains. Since the pulse trains for the two different rates had longest IPIs of the same duration, this implies that the absolute length of the IPI is not as important as the relative IPI in the context of the surrounding pulses. Similar results can be found using electric stimulation (Laback and Majdak, 2008). Two listeners for which comparable data are available showed significantly higher percent correct scores for jittered pulse trains ($k=3/4$) at 1515 pps compared to the scores for periodic pulse trains at 800 pps for an ITD of 600 μs. Nevertheless, there is no reason to assume that long IPIs do not improve ITD sensitivity. However, having only long IPIs is not sufficient to improve ITD sensitivity. Rather, the temporal jitter, which combines both long and short IPIs, seems to be the necessary condition for improved ITD sensitivity.

We modeled the neural response characteristics in order to determine if response changes in the auditory periphery and brainstem might reflect the behavioral changes in ITD sensitivity. The results indicate that jitter increases the synchrony in the neural spike pattern of the ongoing signal. This is especially the case for auditory filters with BFs higher than the center frequency of the stimulus. It is quite likely that the increased synchrony makes it easier for the binaural system to detect an ITD, given that the jitter is synchronized between the two ears.

Modeling the basic operation of MSO neurons, we also showed improved rate-ITD tuning for auditory filters with BFs higher than the center frequency of the stimulus. While the simple MSO model was able to capture some of the expected trends in the data, numerous improvements could be made to the modeling, which might show a greater contrast between the rate-ITD tuning curves for periodic and jittered signals. For example, inclusion of spherical bushy VCN cells, which act as the input to the MSO, may act as monaural coincidence detectors (Carney, 1990). Also, a true physiological model of the MSO could be used (Han and Colburn, 1993), particularly one that includes elements that improve timing aspects, such as the inclusion of dendrites (Agmon-Snir *et al.*, 1998). VCN bushy cells and MSO principle neurons contain low-threshold potassium channels, which are thought to play a role in coincidence detection (Manis and Marx, 1991; Smith, 1995). High-rate pulse trains, which would produce a relatively constant synaptic input to these neurons, may produce sustained activation of the low-threshold potassium channels, which, if included in the model, would suppress neuron repolarization and block firing (Colburn *et al.*, 2008). Also, inclusion of inhibitory effects in the VCN (Burkard and Palmer, 1997) or at higher centers like the inferior colliculus (Smith and Delgutte, 2008) may also help the use of models to understand the jitter effect.

The modeling results provide an explanation for the jitter effect in terms of increased synchrony of the neural response at the level of the AN, which is not inconsistent with recent ITD sensitivity measurements in CI listeners by van Hoesel (2008). This explanation is somewhat different from the hypothesis proposed by Hafter and Buell (1990) that the recovery from binaural adaptation induced by inserting a change (trigger) in a pulse train is an active process involving some kind of change detector. Hafter and Buell (1990) reported that other types of changes applied to the ongoing part of a pulse train, such as the insertion of short trigger signals (either monotic or diotic), cause recovery. The question arises if the effect of these changes could also be explained by an increase in synchrony. The answer is probably no, since a monotic or a diotic trigger signal in the spectral frequency region of the pulse train could disrupt the ITD. This is because the changes in the neural firing pattern would be unassociated with the ITD. This seems to imply that the recovery effect observed by Hafter and Buell (1990) with those triggers is mediated via another mechanism, which may involve a true change detector. However, our modeling results do not necessarily rule out the restarting explanation of Hafter and Buell (1990), and further work needs to be done to investigate the underlying mechanisms of recovery from binaural adaptation.

An interesting aspect of the data obtained in this study, which was also seen for electric hearing in Laback and Majdak (2008), is that binaural jitter resolves the ambiguity in the ongoing ITD cue that occurs whenever the ITD exceeds one-quarter of the IPI. Majdak *et al.* (2006) showed that CI listeners lateralize periodic pulse trains to the wrong (lagging) side for fine-structure ITDs exceeding one-half of the IPI. However, in our study, for jittered pulse trains, listeners lateralize to the correct side, even for ITDs that approach or exceed one-half of the IPI. For example, the NH listeners lateralized jittered 1200-pps pulse trains (IPI $=833$ μs) with a 400-μs ITD, which is about one-half IPI, to the correct (leading) side in nearly all the trials. The CI listeners in Laback and Majdak (2008) correctly lateralized jittered pulse trains at rates from 800 to 1515 pps with ITDs falling within the range of one-quarter to one IPI. There are at least two possible explanations of how binaural jitter could resolve the ITD ambiguity. First, the auditory system could process and analyze the jittered pulse trains as a temporal structure, thus integrating information across time. For a pulse train with an ITD of one-half of the IPI, the classical cross-correlation model of binaural interaction (e.g., Colburn, 1977) predicts an ambiguous pair of peaks in case of a periodic pulse train. However, in the case of a jittered pulse train, the "wrong" peak disappears and only the peak corresponding to the correct ITD remains. Second, the auditory system could pick out interaural pulse pairs with a large IPI to adjacent pairs. This corresponds to a so-called multiple looks model (Viemeister and Wakefield, 1991), where the auditory system stores samples or "looks" of the signal in memory and accesses and processes them selectively.

Finally, based on these findings, we would like to reconsider the interpretation of experiments on the ITD sensitivity to high-frequency bandpass-filtered white noise. In particular, Bernstein and Trahiotis (1994) showed that the ITD JND for bandpass-filtered noise centered at 4 kHz is independent of noise bandwidth up to a bandwidth of at least 800 Hz. The mean envelope rate in filtered noise corresponds to about 64% of the bandwidth (Rice, 1953). Thus, the 800-Hz bandwidth corresponds to a modulation rate of 512 Hz. For sinusoidal AM tones, two-tone complexes, or transposed tones, ITD JNDs could not be measured at modulation rates of 512 Hz or greater (Bernstein and Trahiotis, 1994, 2002). In order to explain the comparatively high ITD sensitivity for the noise, Bernstein and Trahiotis (1994) suggested that the listeners may shift their attention to lower-frequency "internal filters" or critical bands, which would result in a narrower critical bandwidth and consequently a lower rate of envelope fluctuation. Bernstein and Trahiotis (1994) also noted that this strategy reduces the rate of envelope fluctuation with no loss in depth of modulation. In light of the results presented in this study, an alternative explanation for the comparatively high sensitivity to filtered noise is the temporal jitter in the envelope.[5] In other words, we propose that it is the random temporal variation in envelope maxima and minima that causes the high ITD sensitivity for noise rather than the strategy of down-shifting the internal filters. Note that down-shifting of filters cannot explain our results as our pulse trains had a constant spectral bandwidth for all amounts of jitter, including the periodic condition. Naturally, it is also possible that the variation in the amplitudes of the envelope maxima is a relevant factor in case of filtered noise. Due to the similar performance for jittered pulse trains and filtered noise in our study, we assume that the temporal variation is the more important factor.

ACKNOWLEDGMENTS

We would like to thank Mr. Michael Mihocic for running experiments and our listeners. We would like to thank the associate editor Dr. Richard Freyman and two anonymous reviewers for numerous improvements to this work. We would like to thank Dr. Brian Moore, Dr. Zachery Smith, Dr. Andrew Brughera, Dr. Laurel Carney, and Dr. Philip Joris for useful discussions about the binaural jitter phenomenon. We would like to thank Dr. Raymond Meddis for help using his model. We would like to particularly thank Dr. Bertrand Delgutte and Dr. Kenneth Hancock for helping us to understand the pertinent physiology. This study was funded in part by the Austrian Science Fund (FWF Project No. P18401-B15).

[1]The $k=1/3$ value was really $516/1667=0.31$. We intended to use $k=1/4$ but there was a mistake in the experimental program.

[2]JNDs were estimated from a maximum-likelihood cumulative Gaussian fit to the P_c data using PSIGNIFIT version 2.5.41 (see http://bootstrap-software.org/psignifit/, Last viewed 6/18/09), a software package for fitting psychometric functions to psychophysical data (Wichmann and Hill, 2001a, 2001b). The function used was a Weibull function.

[3]Due to the extremely large amount of time needed to compute 30 000 AN firing patterns (30 AN fibers $\times$100 MSO measurements) per condition, we assumed that 30 AN fiber PSTHs randomly chosen from a pool of 100 were sufficient to represent variance of 30 000 PSTHs.

[4]Refractory times of 2 and 4 ms were also tried. The different refractory times resulted in the same basic trends in the data.

[5]Perceptually, periodic pulse trains have a tonal quality. Jitter introduces a noisy or scratchy quality to the pulse trains. The physical and perceptual qualities of temporally-jittered pulse trains and noise were summarized in Pierce *et al.* (1977). In that study, it is stated that "… the central limit theorem tells us that at high enough rates, for which many pulses do overlap, the [jittered pulse train] approaches Gaussian noise." Hence, the ITD sensitivity for jittered pulse trains being bounded by the performance for noise seems consistent with the fact that jittered pulse trains become physically and perceptually similar to noises.

Agmon-Snir, H., Carr, C. E., and Rinzel, J. (**1998**). "The role of dendrites in auditory coincidence detection," Nature (London) **393**, 268–272.

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

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

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

Blauert, J. (**1981**). "Lateralization of jittered tones," J. Acoust. Soc. Am. **70**, 694–698.

Burkard, R., and Palmer, A. R. (**1997**). "Responses of chopper units in the ventral cochlear nucleus of the anaesthetised guinea pig to clicks-in-noise and click trains," Hear. Res. **110**, 234–250.

Carney, L. H. (**1990**). "Sensitivities of cells in anteroventral cochlear nucleus of cat to spatiotemporal discharge patterns across primary afferents," J. Neurophysiol. **64**, 437–456.

Colburn, H. S. (**1977**). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise," J. Acoust. Soc. Am. **61**, 525–533.

Colburn, H. S., Chung, Y., Zhou, Y., and Brughera, A. (**2008**). "Models of brainstem responses to bilateral electrical stimulation," J. Assoc. Res. Otolaryngol. **10**, 91–110.

Dye, R. H., Jr., and Hafter, E. R. (**1984**). "The effects of intensity on the detection of interaural differences of time in high-frequency trains of clicks," J. Acoust. Soc. Am. **75**, 1593–1598.

Hafter, E. R. (**1997**). "Binaural adaptation and the effectiveness of a stimulus beyond its onset," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Earlbaum, Hillsdale, NJ), pp. 211–232.

Hafter, E. R., and Buell, T. N. (**1990**). "Restarting the adapted binaural system," J. Acoust. Soc. Am. **88**, 806–812.

Hafter, E. R., and Dye, R. H., Jr. (**1983**). "Detection of interaural differences of time in trains of high-frequency clicks as a function of interclick interval and number," J. Acoust. Soc. Am. **73**, 644–651.

Han, Y., and Colburn, H. S. (**1993**). "Point-neuron model for binaural interaction in MSO," Hear. Res. **68**, 115–130.

Henning, G. B. (**1974**). "Detectability of interaural delay in high-frequency complex waveforms," J. Acoust. Soc. Am. **55**, 84–90.

Henning, G. B. (**1980**). "Some observations on the lateralization of complex waveforms," J. Acoust. Soc. Am. **68**, 446–454.

Huber, A., Linder, T., Ferrazzini, M., Schmid, S., Dillier, N., Stoeckli, S., and Fisch, U. (**2001**). "Intraoperative assessment of stapes movement," Ann. Otol. Rhinol. Laryngol. **110**, 31–35.

Joris, P. X., Louage, D. H., Cardoen, L., and van der Heijden, M. (**2006**). "Correlation index: A new metric to quantify temporal coding," Hear. Res. **216–217**, 19–30.

Laback, B., and Majdak, P. (**2008**). "Binaural jitter improves interaural time-difference sensitivity of cochlear implantees at high pulse rates," Proc. Natl. Acad. Sci. U.S.A. **105**, 814–817.

Laback, B., Majdak, P., and Baumgartner, W. D. (**2007**). "Lateralization discrimination of interaural time delays in four-pulse sequences in electric and acoustic hearing," J. Acoust. Soc. Am. **121**, 2182–2191.

Lopez-Poveda, E. A., and Meddis, R. (**2001**). "A human nonlinear cochlear filterbank," J. Acoust. Soc. Am. **110**, 3107–3118.

Majdak, P., and Laback, B. (**2009**). "Effects of center frequency and rate on the sensitivity to interaural delay in high-frequency clicks," J. Acoust. Soc. Am. **125**, 3903–3913.

Majdak, P., Laback, B., and Baumgartner, W. D. (**2006**). "Effects of interaural time differences in fine structure and envelope on lateral discrimination in electric hearing," J. Acoust. Soc. Am. **120**, 2190–2201.

Manis, P. B., and Marx, S. O. (**1991**). "Outward currents in isolated ventral cochlear nucleus neurons," J. Neurosci. **11**, 2865–2880.

Meddis, R. (**2006**). "Auditory-nerve first-spike latency and auditory absolute threshold: A computer model," J. Acoust. Soc. Am. **119**, 406–417.

Moore, B. C., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.

Nordmark, J. O. (**1976**). "Binaural time discrimination," J. Acoust. Soc. Am. **60**, 870–880.

Pierce, J. R., Lipes, R., and Cheetham, C. (**1977**). "Uncertainty concerning the direct use of time information in hearing: Place clues in white-spectra stimuli," J. Acoust. Soc. Am. **61**, 1609–1621.

Rice, S. O. (**1953**). "Mathematical analysis of random noise," in *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York), pp. 133–294.

Saberi, K., and Perrott, D. R. (**1995**). "Lateralization of trains of clicks with opposing onset and ongoing interaural delays," Acustica **81**, 272–275.

Scott, L. L., Hage, T. A., and Golding, N. L. (**2007**). "Weak action potential backpropagation is associated with high-frequency axonal firing capability in principal neurons of the gerbil medial superior olive," J. Physiol. **583**, 647–661.

Scott, L. L., Mathews, P. J., and Golding, N. L. (**2005**). "Posthearing developmental refinement of temporal processing in principal neurons of the medial superior olive," J. Neurosci. **25**, 7887–7895.

Smith, P. H. (**1995**). "Structural and functional differences distinguish principal from nonprincipal cells in the guinea pig MSO slice," J. Neurophysiol. **73**, 1653–1667.

Smith, Z. M., and Delgutte, B. (**2008**). "Sensitivity of inferior colliculus neurons to interaural time differences in the envelope versus the fine structure with bilateral cochlear implants," J. Neurophysiol. **99**, 2390–2407.

Smith, P. H., Joris, P. X., and Yin, T. C. (**1993**). "Projections of physiologically characterized spherical bushy cell axons from the cochlear nucleus of the cat: Evidence for delay lines to the medial superior olive," J. Comp. Neurol. **331**, 245–260.

Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.

van Hoesel, R. J. M. (**2007**). "Sensitivity to binaural timing in bilateral cochlear implant users," J. Acoust. Soc. Am. **121**, 2192–2206.

van Hoesel, R. J. M. (**2008**). "Observer weighting of level and timing cues in bilateral cochlear implant users," J. Acoust. Soc. Am. **124**, 3861–3872.
Viemeister, N. F., and Wakefield, G. H. (**1991**). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–865.
Wichmann, F. A., and Hill, N. J. (**2001a**). "The psychometric function: I. Fitting, sampling, and goodness of fit," Percept. Psychophys. **63**, 1293–1313.
Wichmann, F. A., and Hill, N. J. (**2001b**). "The psychometric function: II. Bootstrap-based confidence intervals and sampling," Percept. Psychophys. **63**, 1314–1329.

Role of binaural hearing in speech intelligibility and spatial release from masking using vocoded speech

Soha N. Garadat,[a] Ruth Y. Litovsky,[b] and Gongqiang Yu[c]
Waisman Center, University of Wisconsin, 1500 Highland Avenue, Madison, Wisconsin 53705

Fan-Gang Zeng
University of California-Irvine, 364 Med Surge II, Irvine, California 92697

(Received 11 October 2007; revised 31 August 2009; accepted 31 August 2009)

A cochlear implant vocoder was used to evaluate relative contributions of spectral and binaural temporal fine-structure cues to speech intelligibility. In Study I, stimuli were vocoded, and then convolved through head related transfer functions (HRTFs) to remove speech temporal fine structure but preserve the binaural temporal fine-structure cues. In Study II, the order of processing was reversed to remove both speech and binaural temporal fine-structure cues. Speech reception thresholds (SRTs) were measured adaptively in quiet, and with interfering speech, for unprocessed and vocoded speech (16, 8, and 4 frequency bands), under binaural or monaural (right-ear) conditions. Under binaural conditions, as the number of bands decreased, SRTs increased. With decreasing number of frequency bands, greater benefit from spatial separation of target and interferer was observed, especially in the 8-band condition. The present results demonstrate a strong role of the binaural cues in spectrally degraded speech, when the target and interfering speech are more likely to be confused. The nearly normal binaural benefits under present simulation conditions and the lack of order of processing effect further suggest that preservation of binaural cues is likely to improve performance in bilaterally implanted recipients.
 [DOI: 10.1121/1.3238242]

PACS number(s): 43.66.Pn, 43.66.Qp, 43.66.Ts, 43.71.Ky [RLF] Pages: 2522–2535

I. INTRODUCTION

Cochlear implants (CIs) have been highly successful at providing hearing to profoundly deaf individuals. As a result of continual progress made in this advanced technology, auditory perception in recipients has improved significantly in the past few decades. Today, most CI users are able to perform well in quiet listening situations. However, their performance deteriorates considerably in the presence of background noise and competing speech (Skinner *et al.*, 1994; Muller-Deile *et al.*, 1995; Battmer *et al.*, 1997; Stickney *et al.*, 2004). Numerous studies that focus on performance in unilateral CI users have attempted to identify some of the factors that can account for this deterioration, including the role of speech coding strategies and number of frequency bands (e.g., Gantz *et al.*, 1988; Waltzman *et al.*, 1992; Dorman and Loizou, 1997; Friesen *et al.*, 2001; Stickney *et al.*, 2004). In an alternative approach, bilateral CIs have been provided to a growing number of recipients, with the hope that stimulation of both ears will lead to improved performance in difficult listening situations. Results to date suggest that many bilateral CI users perform better at understanding speech in adverse listening conditions when using two CIs compared with a single CI (e.g., Schleich *et al.*, 2004; Iwaki *et al.*, 2004; Litovsky *et al.*, 2004; 2006; 2009; Tyler *et al.*, 2002). However, despite this improved performance, bilateral CI users are still considerably challenged in dynamic listening situations. In addition, there remains a gap in performance between bilateral CI users and normal hearing listeners (NHLs). The reasons for the gap remain to be understood.

When addressing the deficit in speech intelligibility that is experienced by CI users in presence of noise or competing speech, the complexity of the everyday auditory scene should be considered alongside the possibility that performance is limited by signal processing in the prosthetic devices. In real-world listening, the signal and unwanted "competing" sounds may overlap spectrally and temporally, as well as spatially. Often, there also exists acoustic variability in the auditory environments that may result in increased similarity between a target sound and a competing source, rendering extraction of the target signal rather difficult. The difficulty associated with source segregation under such conditions is often attributed to informational masking (Neff, 1995; Brungart, 2001; Kidd *et al.*, 2002). Although the effects of informational masking can be decreased by introducing dissimilarity between the target and interferer (Durlach *et al.*, 2003), this approach might not be realistically feasible in many listening situations due to the unpredictability of the auditory environments.

Overcoming informational masking can be achieved if listeners have access to a variety of other auditory cues. For example, NHLs exploit spectral (Assmann and Summerfield, 1990, 1994; Bird and Darwin, 1998; Vliegen and Oxhenham, 1999) as well as temporal (Tyler *et al.*, 1982; Buss and Flo-

[a]Present address: Kresge Hearing Research Institute, University of Michigan-Ann Arbor, MI.
[b]Author to whom correspondence should be addressed. Electronic mail: litovsky@waisman.wisc.edu
[c]Present address: University of Connecticut, Farmington, CT.

rentine, 1985; Bacon *et al.*, 1998; Summers and Molis, 2004) cues to segregate overlapping and competing auditory streams. It is also well known that NHLs can take advantage of spatial cues to segregate speech from competing sounds. This is manifested in as much as a 12 dB improvement in speech reception thresholds (SRTs) when target speech and competitors are spatially separated compared with situations in which they are co-located. This benefit is known as spatial release from masking (SRM), and is an effect that has been studied extensively in NHLs (Freyman *et al.*, 1999; Arbogast *et al.*, 2005; Hawley *et al.*, 1999; 2004; Drullman and Bronkhorst, 2000; Litovsky, 2005).

Spatial cues appear to become more prominent under conditions in which informational masking is relatively large (Kidd *et al.*, 1998; Arbogast *et al.*, 2002). This suggests that spatial hearing plays a crucial role in helping listeners to overcome informational masking. Given the growing number of bilateral CI users, the extent to which spatial cues can be made available to these listeners is a timely question with regard to addressing the gap in performance noted above. The contribution of spatial cues can be explored in these individuals by controlling the inputs to the two ears and comparing performance under bilateral vs unilateral listening modes. A recent study by Loizou *et al.* (2009) has shown that, compared with NHLs, bilateral users are less capable of taking advantage of binaural cues for source segregation, in particular, under conditions of informational masking. This may be due to the fact that CIs have limited spectral resolution (Freisen *et al.*, 2001) and ineffective encoding of F0 information (Stickney *et al.*, 2007). The novelty of the study of Loizou *et al.* (2009) lies in the tighter stimulus control utilized by presenting binaural stimuli directly to the CI users' processors, with spatially appropriate stimuli that were convolved with head related transfer functions (HRTFs).

Limitations in performance of participants in the study of Loizou *et al.* (2009) are of interest here, as they may have arisen from two factors that are highly difficult to control in CI users. One factor is the lack of obligatory coordination between specific pairs of electrodes across the two ears, which would have reduced the extent to which binaural cues could be preserved with fidelity upon reaching the brainstem. A second issue arises when participants whose auditory system has undergone periods of auditory deprivation are tested. Disruptions in the neural processing mechanisms are likely to be present and to contribute to variability in performance within the population of CI users, leading to difficulty in identifying and understanding mechanisms involved in the processes of binaural cues under complex listening conditions.

CI vocoders can offer a powerful tool for investigating effects of CI signal processing independently of other confounds inherent in cochlear implantation. In the current study, a CI vocoder was utilized to investigate whether limitations in performance on spatial auditory tasks that are observed in bilateral CI users are due to the signal processing itself. One of the main issues addressed in the present study is whether CI users are susceptible to informational masking that is borne out of crude signal processing in their prosthetic devices. This issue was investigated by using testing conditions that represent simple but realistic everyday listening situations, yet at the same time in which informational masking in the non-CI conditions may be small. Spondaic target words were presented in the presence of sentences, a combination of target and interferer that deliberately creates relatively easy testing conditions. This approach enabled a systematic examination of a number of critical factors related to speech intelligibility in adverse listening conditions, akin to those that occur with CI processors when a limited number of frequency bands are available. Specifically, speech intelligibility and SRM were evaluated using spectrally degraded stimuli, under binaural and monaural listening conditions. Of a particular interest in this study was the extent to which CI signal processing might impact the role of binaural hearing in providing benefits on measures of speech intelligibility and SRM.

Listening conditions in this study utilized "virtual space" techniques (e.g., Hawley *et al.*, 2004; Loizou *et al.*, 2009) such that all acoustic stimuli were convolved with HRTFs[1] to introduce more realistic, perceptually spatialized and separated target and competing stimuli. By controlling the stage at which stimuli were convolved with HRTFs, effects of signal processing and CI vocoding can be examined independent of the potential loss of binaural cues. Given that one of the future goals in bilateral CIs is to design and provide systems that capture and mimic the way that acoustic information is transmitted in NHLs, the present study could shed light on factors that could potentially enhance vs impair outcomes for effects due to binaural squelch, binaural summation, and the head-shadow effect.

II. STUDY I

In this study, conditions that are more idealized relative to true CI listening were examined by first processing the speech stimuli through the vocoders and subsequently convolving the output through the HRTFs. This approach is akin to a situation in which a NHL is presented with spectrally degraded stimuli through loudspeakers in a room, an approach that has previously been used to investigate effects of spectral degradation on speech perception but without considering effects of binaural hearing and/or spatial cues (e.g., Shannon *et al.*, 2002; Başkent and Shannon, 2007). The current study was designed to preserve as many cues as possible that would be naturally available to listeners for SRM. These include cues that are known to be available to bilateral CI users to some extent, such as head shadow and envelope interaural time differences (van Hoesel, 2004). In addition, we could preserve cues that contribute to spatialized percepts through temporal fine structure, an important binaural cue that is lost in CI processing. While the original speech fine structure in any band has been replaced with a tone, with the idealized order of processing applied here, the new fine structure is filtered through the HRTFs and thus contains the acoustic cues that are used for spatialization.

By preserving the fine-structure cues, it was assumed that there should be sufficient spatial information to acquire the classic release from masking for spectrally degraded

TABLE I. List of cutoff frequencies.

	16-band			8-band			4-band		
Band	*Lf*	*Cf*	*Hf*	*Lf*	*Cf*	*Hf*	*Lf*	*Cf*	*Hf*
1	300	350	400	300	411	521	300	574	848
2	400	460.5	521	521	686	848	848	1445	2042
3	521	595	669	848	1089	1330	2042	3341	4640
4	669	758.5	848	1330	1686	2042	4640	7470	10300
5	848	957	1066	2042	2566	3091			
6	1066	1198	1330	3091	3866	4640			
7	1330	1490.5	1651	4640	5784	6927			
8	1651	1845.5	2042	6927	8613	10300			
9	2042	2279	2516						
10	2516	2803.5	3091						
11	3091	3441	3791						
12	3791	4215.5	4640						
13	4640	5156.5	5673						
14	5673	6300	6927						
15	6927	7688.5	8450						
16	8450	9375	10300						

stimuli; hence, informational masking that is created by signal processing can be evaluated with limited confounds.

A. Material and methods

1. Listeners

Nine NHLs (three male, six female; age range 19–25 years) participated. All subjects were native speakers of English and had pure tone thresholds better than 15 dB hearing loss at octave frequencies ranging from 250–8000 Hz. Participants signed a consent form approved by University of Wisconsin-Madison Institutional Review Board and were paid for their participation. Testing was conducted in five two-hour sessions.

2. Signal processing

Speech signals with a bandwidth between 300 and 10300 Hz^2 were bandpass filtered into 4, or 8, or 16 contiguous frequency bands (see Table I) by sixth-order Butterworth filters using a MATLAB software simulation of CI signal processing strategies (e.g., Shannon *et al.*, 1995). Briefly, the envelope was extracted from each band by full-wave rectification and low-pass filtering at 50 Hz with a second order Butterworth filter. The extracted envelope was used to amplitude modulate a sinusoidal carrier at the band's central frequency followed by the same bandpass filter as the analysis filter to remove spectral splatter. All bands were summed and then convolved with HRTFs (Gardner and Martin, 1994) to create perceptually spatialized and virtually separated target and interferers. For each stimulus (target or interferer), the carrier tones in the right and left ears were in phase. The phase relationship between the carrier tones for target and interferer waveforms was arbitrary. Target and interfering stimuli were then summed and presented to the listeners through headphones (Sennheiser HD 580) under binaural and monaural (right-ear) conditions. In the vocoded speech conditions, target and interfering sentences were processed in the same manner.

3. Stimuli materials and virtual spatial configuration

Target stimuli consisted of a closed set of 25 spondees recorded in our laboratory with a male-talker and presented in quiet as well as in the presence of competing speech. The interferer stimuli were sentences from the Harvard IEEE corpus (Rothauser *et al.*, 1969) recorded with a different male talker than the target. Thirty sentences were strung together, and segments were randomly chosen and played for 6 s per trial. The target words began approximately 1.5 s after the onset of the competing sentence. On each trial the 25 spondees were visually presented to the subjects on a computer monitor. Subjects were instructed to respond by using a mouse button to select the appropriate target word. Feedback was provided following each response by flanking the correct stimuli on the computer screen in front of the listener.

Given that all stimuli were convolved through HRTFs to enable virtual spatial separation of target and interfering speech, data were collected for each subject using the following location combinations: (1) quiet: target at 0° azimuth and no interferer, (2) front: target and interferer both at 0° azimuth, (3) right: target at 0° azimuth and interferer at 90° azimuth, and (4) left: target at 0° azimuth and interferer at −90° azimuth.

4. Stimulus levels and threshold estimation

All stimuli were calibrated using an artificial ear coupler (AEC101 IEC 318, Larson Davis). Calibration was conducted after stimuli were convolved through the HRTFs. Stimulus levels were set based on calibration for token sentences from the speech corpus presented from the simulated front condition. The level of the interferer was fixed at 60 dB sound pressure level (SPL); thus for the front condition, interferer levels were set to 60 dB SPL in each ear. For nonfront conditions, interferer levels were 60 dB SPL for the ear ipsilateral to the interferers, and change in signal-to-noise ratio (SNR) represents the change in target level relative to interferer level at the ipsilateral ear. The level varied natu-

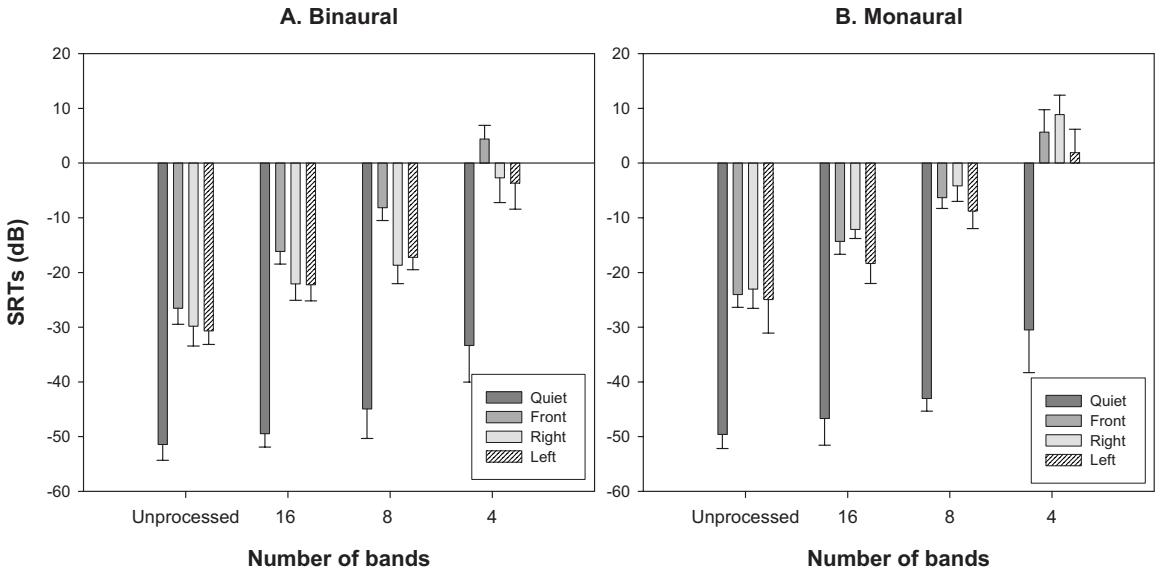


FIG. 1. Average SRTs (+1 SD) in dB are shown for all spatial conditions tested relative to interferer level (60 dB). Left panel represent binaural data and right panel represents monaural (right-ear) data. Within each panel, SRTs are plotted for the different interferer conditions as a function of frequency band conditions.

rally with the HRTF at the contralateral side to create a head-shadow effect. The level of the target was varied adaptively using an algorithm that targets the 79.4 point on the psychometric function (Levitt, 1971). The target level was initially 65 dB SPL and was decremented by 8 dB following each correct response. After the first incorrect response, a modified adaptive 3-down/1-up algorithm was used in which the step size was halved after each reversal, with the minimum step size set to 2 dB. If the same step size was used twice in a row in the same direction, the next step size was doubled in value. Testing was terminated following eight reversals.

SRTs were estimated from the adaptive tracks by using a constrained maximum-likelihood method of parameter estimation (MLE), which has been described by Wichmann and Hill (2001a, 2001b). Based on this method, data from each experimental run for each participant were fitted to a logistic function and thresholds were calculated by taking the level of the target at a specific probability level. This approach has been shown to yield comparable results to the well-known approach in which SRT is defined as the average of levels at which reversals occur. However, the MLE approach has the advantage, with this stimulus corpus and adaptive tracking method, of producing smaller group variance (Litovsky, 2005).

5. Procedure and training

Data collection was conducted in blocks with the number of frequency bands (unprocessed, 16, 8, and 4) fixed. To ensure familiarity with the task, participants completed the unprocessed conditions first. Subsequently, all other blocks (16, 8, and 4) were presented in a random order generated with a different seed for each subject. Within a block, all other conditions were randomized. Prior to each testing block with vocoded stimuli, subjects received additional listening exposures to familiarize them with the quality of the speech they were about to hear in the upcoming blocked condition. During these vocoded exposure periods, four SRTs (two in quiet, and two with front interferer) were collected; these SRTs were excluded from the main analyses. After data completion, all conditions were re-randomized and a second set of data was collected based on the assumption that with more exposure to vocoded speech, listeners' performance would be more stable. This second set of data was used in the analyses and reported in this paper. However, statistical analyses comparing the two sets of data revealed that learning effects occurred only in the 4-band conditions.

B. Results

1. SRTs

SRTs were obtained using the MLE procedure described above and were normalized relative to interferer level. These data are displayed in Fig. 1 as a function of number of frequency bands under binaural and monaural conditions. Two-way repeated measure analyses of variance (ANOVAS) on SRTs were conducted for listening mode (binaural and monaural) and number of frequency bands (unprocessed, 16, 8, and 4); these analyses were conducted separately for each interferer condition (quiet, front, right, and left). A significant main effect of listening mode was found such that binaural SRTs were lower than monaural SRTs in all conditions; quiet $[F(1,8)=9.874,\ p<0.05]$, front $[F(1,8)=14.752,\ p<0.01]$, right $[F(1,8)=77.763,\ p<0.0001]$, and left $[F(1,8)=42.205,\ p<0.0001]$.

Significant main effects of number of bands were also observed for all conditions; quiet $[F(3,8)=53.243,\ p<0.0001]$, front $[F(3,8)=571.718,\ p<0.0001]$, right $[F(3,8)=298.448,\ p<0.0001]$, and left $[F(3,8)=2364.374,\ p<0.0001]$ SRTs. The lack of interactions with listening mode suggests that the effect of number of bands applies to binaural and monaural listening modes. *Post-hoc* Scheffe's tests revealed that, in quiet, SRTs for the unprocessed condition were comparable to those in the 16-band condition but lower (better performance) than those in the 8- and 4-band

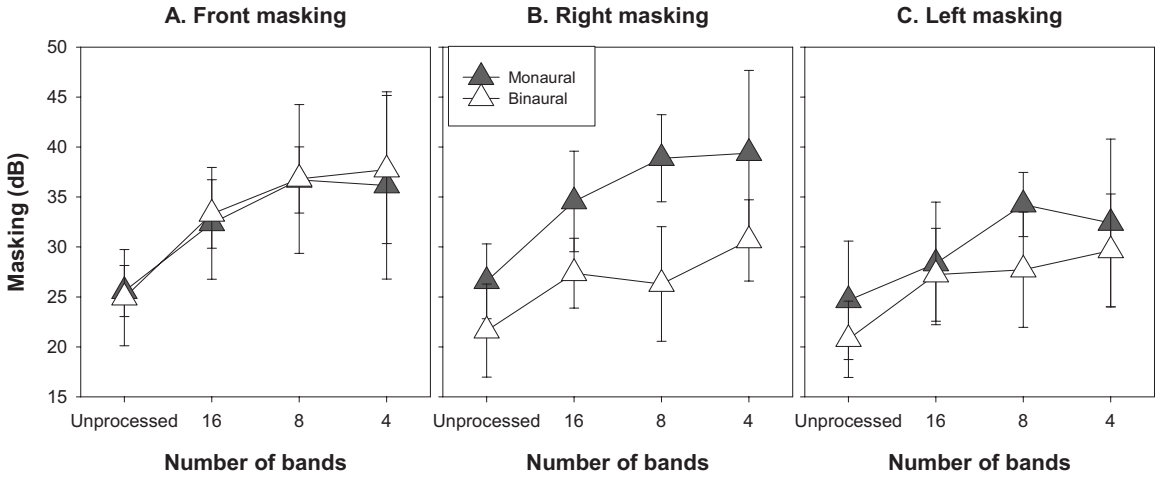


FIG. 2. Masking values (±1 SD) are plotted for the monaural (filled symbols) and binaural (unfilled symbols) as a function of the different frequency bands and shown for the front (panel A), right (panel B), and left (panel C) interferer conditions.

conditions ($p<0.001$). However, in the presence of a speech interferer, the improvement in SRTs continued with further increases in number of frequency bands ($p<0.0001$). Specifically, SRTs for the unprocessed conditions were lower than those in the 16-, 8-, and 4-band conditions. In addition, SRTs for the 16-band conditions were lower than those in the 8- and 4-band conditions, with lower SRTs in the 8-band than those in the 4-band conditions.

2. Masking

In this study, masking was defined as the absolute change in SRTs when interferer stimuli were present compared with the quiet condition. Masking values for the front, right, and left, respectively, were computed as ($SRT_{front}-SRT_{quiet}$), ($SRT_{right}-SRT_{quiet}$), and ($SRT_{left}-SRT_{quiet}$). These masking values, shown in Fig. 2, were subjected to two-way repeated measures ANOVAs for listening mode (binaural, monaural) and number of frequency bands (unprocessed, 16, 8, and 4) as described above for SRTs.

A main effect of listening mode was not found for front masking, indicating comparable amount of masking for the binaural and monaural conditions. However, a main effect of listening mode was obtained for right [$F(1,8)=88.280$, $p<0.0001$] and left [$F(1,8)=13.346$, $p<0.01$] maskings, such that the amount of masking was greater in the monaural than in the binaural conditions. These results suggest that binaural listening provides mechanisms for reduction in masking that are not available in the single-ear listening mode. In addition, a main effect of number of frequency bands was obtained for front [$F(3,8)=20.502$, $p<0.0001$], right [$F(3,8)=14.511$, $p<0.0001$], and left [$F(3,8)=6.944$, $p<0.005$] masking. Scheffe's *post-hoc* analyses revealed that the amount of masking was significantly smaller in the unprocessed condition than the 16-, 8-, and 4-band conditions ($p<0.01$), suggesting that spectrally degraded speech is more susceptible to masking than natural speech. Finally, differences in masking were not statistically significant across the three spectrally degraded conditions; this finding occurred in all three spatial masker configurations: front, right, and left.

3. Spatial release from masking

Figure 3 summarizes the findings for SRM which was computed for two spatial configurations: right ($Masking_{front}-Masking_{right}$) and left ($Masking_{front}-Masking_{left}$). These data were subjected to two-way repeated measures ANOVAs for listening mode (binaural and monaural), and number of frequency bands (unprocessed, 16, 8, and 4); separate analyses were conducted for the right and left SRM values. A main effect of listening mode suggested that SRM was larger in the binaural than in the monaural conditions (right-ear) for both right [$F(1,8)=51.317$, $p<0.0001$] and left [$F(1,8)=24.700$, $p<0.005$] interferer configurations.

A main effect of number of frequency bands was not found for the right spatial configuration but was obtained for the left configuration [$F(1,8)=3.424$, $p<0.05$]. Scheffe's *post-hoc* analysis revealed that, in comparison with the unprocessed condition, the amount of SRM was greater for spectrally degraded conditions: 16-band ($p<0.05$), 8-band ($p<0.005$), and 4-band ($p<0.001$). Differences in SRM were not statistically significant across the different spectrally degraded conditions.

A significant interaction of listening mode × number of frequency bands [$F(3,24)=5.116$, $p<0.01$] was found for the right spatial configuration. Scheffe's *post-hoc* analysis showed that, under monaural listening, SRM was statistically comparable for the different spectral conditions. In the binaural conditions, SRM for the unprocessed condition was smaller than that in the 8 ($p<0.0001$) and 4 ($p<0.005$) bands, and comparable to the 16-band condition. In addition, SRM was greater in the 8-band condition compared with 16-band ($p<0.001$) and 4-band ($p<0.05$) conditions. SRM for the 16- and 4-band conditions was comparable.

4. Bilateral effects

Further analyses were conducted in order to facilitate comparisons with studies in bilateral CI users. The variables of interest were head shadow, binaural squelch, and binaural summation (e.g., Muller *et al.*, 2002; Tyler *et al.*, 2003; Schleich *et al.*, 2004; Litovsky *et al.*, 2006). Head shadow in the monaural (right-ear) condition was defined as the advan-

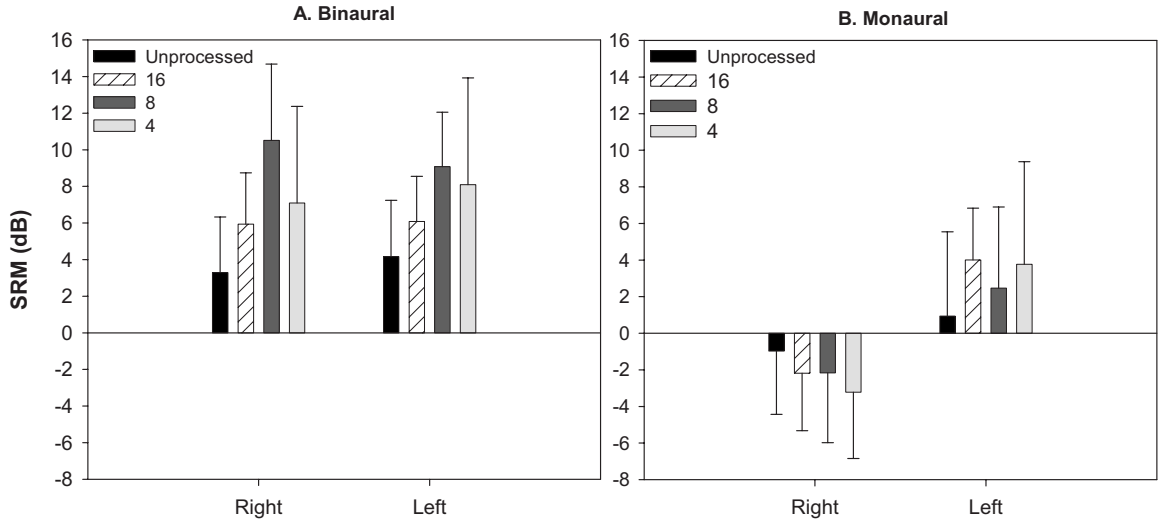


FIG. 3. Average amounts of SRM (+1 SD) are shown for the binaural (A) and monaural (B) listening modes. Within each panel, SRM is compared for the different frequency bands, as a function of interferer location.

tage (reduction in SRT) obtained when the interferer was contralateral versus ipsilateral to the functional ear. It was thus computed as $[\mathrm{SRT(monaural)}_{\mathrm{right}} - \mathrm{SRT(monaural)}_{\mathrm{left}}]$. Binaural squelch describes the advantage obtained as a result of spatial separation between target stimuli and interfering stimuli. These values were obtained for each subject as $[(\mathrm{monaural})_{\mathrm{left}} - (\mathrm{binaural})_{\mathrm{left}}]$. Binaural summation, an advantage that can result from listening to identical stimuli with two ears, was calculated for each subject in two ways; first, by comparing SRTs in the conditions with no interferer $[(\mathrm{monaural})_{\mathrm{quiet}} - (\mathrm{binaural})_{\mathrm{quiet}}]$, and second, by comparing SRTs in the conditions with interferer in the front $[(\mathrm{monaural})_{\mathrm{front}} - (\mathrm{binaural})_{\mathrm{front}}]$.

For each of the four effects listed above, a one-way repeated measures ANOVA was conducted in which the variable of interest was number of frequency bands, including the unprocessed conditions. There were no statistically significant findings for any of the analyses, suggesting that the effects were not dependent on spectral resolution. Data for the vocoded speech were pooled across frequency band conditions and plotted as group means (+1 SD) in Fig. 4. Average values were 5.3 dB for head shadow, 5.9 dB for squelch, and 2.5 and 1.6 dB for binaural summation in quiet and in the presence of the front interferer, respectively. In Fig. 4, the bilateral effects are also plotted for the unprocessed conditions for comparison purposes. Given the lack of a significant main effect of number of spectral bands, the unprocessed and processed conditions were grouped for each condition and were subjected to one-sample *t*-tests (e.g., Schleich *et al.*, 2004). Results revealed that head shadow, squelch, and summation in quiet and in the presence of front interferer were each significantly different than zero ($p < 0.0001$, $p < 0.0001$, $p < 0.01$, $p < 0.01$, respectively).

III. STUDY II

Given the increased robustness of SRM found in the first study when using vocoded speech, the next question addressed here was whether this effect can also be observed in a scenario that more realistically simulates true bilateral CI listening. Therefore, speech stimuli were first convolved through the HRTFs, as would occur in a real world to a person using CIs in the free field; the resulting stimuli were subsequently processed through the vocoder. This study, with the reversed order of signal processing, enabled us to examine whether the directionally dependent cues that are available in the HRTFs are immune to, or distorted by, the CI signal processing in ways that affect benefits from binaural hearing for the spatially separated conditions. Testing was conducted with a second group of listeners, and data from the two studies will be henceforth for conditions that are thought to involve the use of binaural directional cues for source segregation.

A. Material and methods

1. Listeners

Nine NHLs (two male, five female; age range 19–25 years) participated. All subjects were native speakers of Eng-

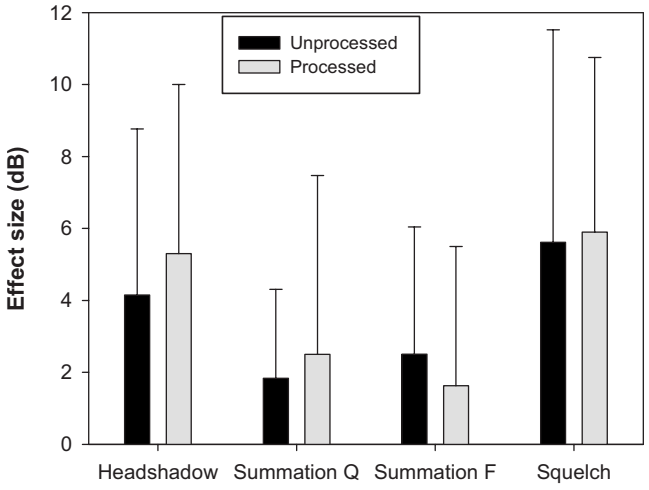


FIG. 4. Group means (+1 SD) are shown for head shadow, binaural squelch, and binaural summation as estimated from the quiet condition and binaural summation as estimated from the condition with interferer in front. Data are plotted for the unprocessed conditions (dark bars) and processed conditions (light bars).

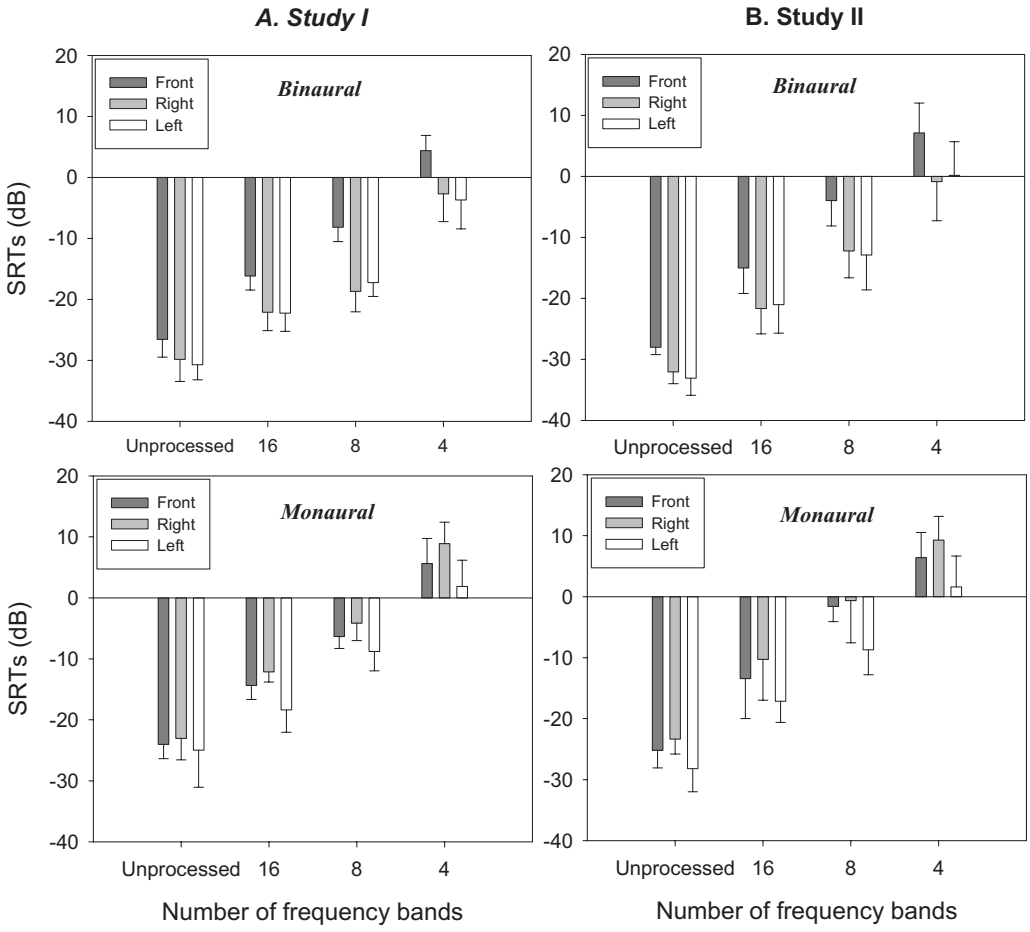


FIG. 5. Average SRTs (+1 SD) in dB are shown for all spatial conditions tested relative to interferer level and displayed for Study I (left column) and Study II (right column). Top panels represent binaural data and bottom panels represent monaural data.

lish and had pure-tone thresholds better than 15 dB HL at octave frequencies ranging from 250–8000 Hz. Participants signed a consent form approved by the University of Wisconsin-Madison Health Sciences Institutional Review Board and were paid for their participation.

2. Signal processing

Similar to Study I, all stimuli were bandpass filtered into 4, or 8, or 16 contiguous frequency bands by sixth-order Butterworth filters with equal bandwidths on a logarithmic scale from 300 to 10300 Hz. Target and interfering waveforms were convolved through HRTFs to create directionally dependent stimuli. Similar to that in Study I, the carrier tones for each of the target and interferer were in phase across the two ears but were not systematically in phase between target and interferer. Stimuli were then digitally mixed and subsequently passed through the CI simulation filters described in detail in Study I. A Tucker-Davis Technologies (TDT) RP2 array processor was used to attenuate the stimuli and to subsequently present them to listeners via TDT system III-hardware (HB7) and headphones (Sennheiser HD 580).

3. Stimuli and procedure

Speech stimuli and the testing apparatus were identical to those described in detail in Study I. Participants were tested on a total of 24 conditions consisting of all combinations of three interferer locations (front, right and left), two listening modes (binaural and monaural/right-ear), and four spectral conditions (unprocessed, 16-, 8- and 4-band conditions). A similar training procedure to that described in the first study was used in this study in order to account for learning effects that might occur when listening to vocoded speech. Given that learning effects only occurred for the 4-band conditions in Study I, only that condition was repeated and included in the data analyses conducted in Study II. Data collection was completed in three two-hour sessions per participant.

B. Results

1. SRTs

SRTs obtained in the two studies are shown in Fig. 5, where results based on the two different orders of processing can be compared. Data were subjected to repeated measures ANOVAs with number of frequency bands (unprocessed, 16, 8, and 4) and interferer conditions (front, right, and left) as the within subject variables and processing order (Study I and Study II) as the between-subject variable. This analysis was conducted separately for binaural and monaural SRTs, in order to directly compare the processing order effects within each listening mode.

There was no significant main effect of processing order for either binaural or monaural SRTs. As was seen when data

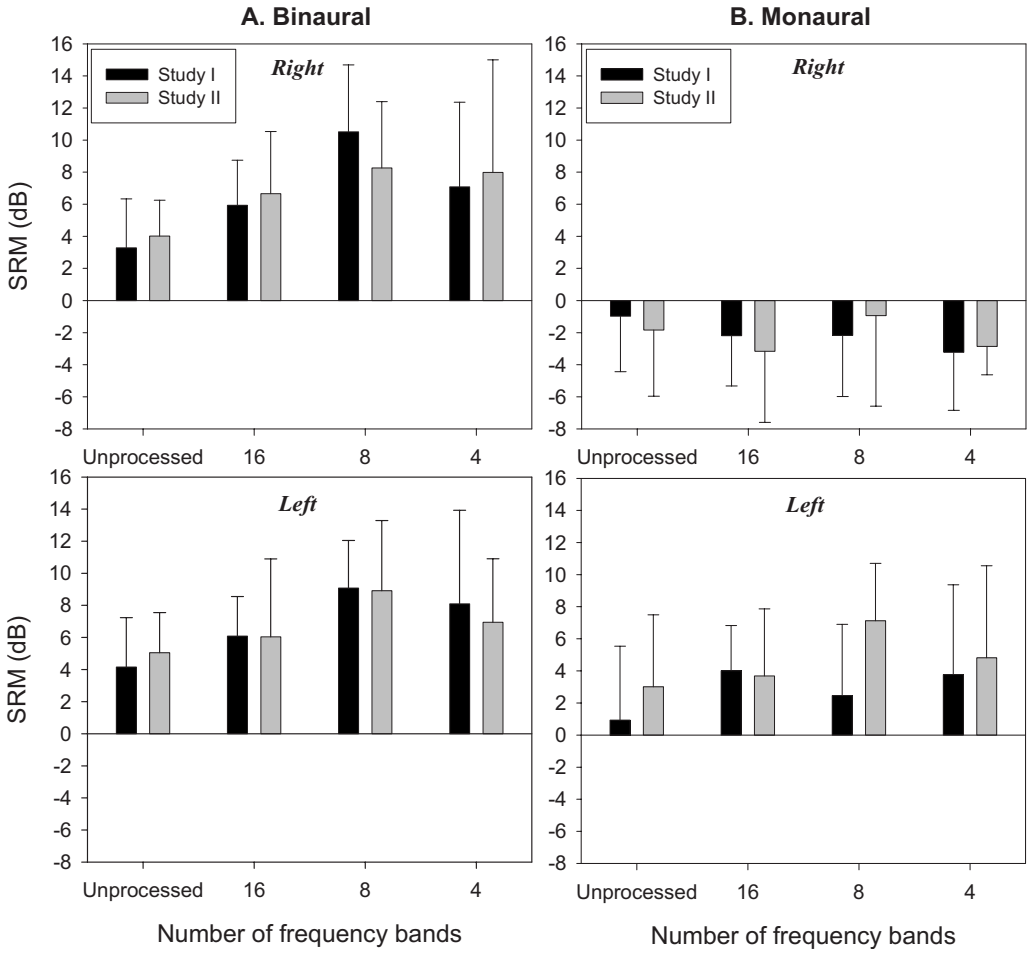


FIG. 6. Average amounts of SRM (+1 SD) are shown for the binaural (left column) and monaural (right column) listening modes. Within each panel, SRM is compared for the two studies across the different frequency bands. Each panel represents different interferer condition.

from Study I were analyzed, a significant main effect of number of frequency bands was found for both binaural $[F(3,16)=341.404,\ p<0.0001]$ and monaural $[F(3,16)=346.285,\ p<0.0001]$ conditions. *Post-hoc* tests were conducted for between-subject effects; all *t*-tests reported in these experiments were corrected for multiple comparisons using the Holm–Bonferroni procedure. Results from the *post-hoc* analyses (significance levels, $p<0.0001$) showed that SRTs for the unprocessed conditions were significantly lower than those in the 16-, 8- and 4-band conditions. In addition, SRTs with 16-bands were lower than those in the 8- and 4-band conditions; SRTs with 8-bands were lower than with 4-bands.

There was a significant main effect of interferer condition for both binaural $[F(2,16)=96.270,\ p<0.0001]$ and monaural $[F(2,16)=86.737,\ p<0.0001]$ SRTs. *Post-hoc* analyses (significance levels, $p<0.0001$) showed that binaural SRTs were higher when the interferer was placed in the front than that in the right or left configurations, the latter two conditions resulted in comparable SRTs. On the other hand, *post-hoc* analyses for the monaural right-ear conditions showed that when the interferer was located in the front, SRTs were higher than when the interferer was on the left but lower than when the interferer was on the right; monaural SRTs for the right were higher than SRTs for the left configuration. Note that the lower monaural SRTs when the interferer was in the front relative to the right occurred despite the calibration procedure that equated the interferer SPLs in the two conditions. The difference in SRTs could have been due to differences in the interferer's spectral profile resulting from frequency-dependent head shadow effects.

A significant interaction was observed for number of frequency bands and processing order in the binaural conditions $[F(3,48)=4.504,\ p<0.01]$ but not in the monaural right-ear conditions. *Post-hoc* tests revealed that for the 8-band conditions, binaural SRTs were lower in Study I compared with Study II. This suggests that when vocoding is conducted on stimuli that have already been filtered with HRTFs, there are likely to be adverse affects on performance.

2. SRM

The approach for deriving SRM values was similar to that used in Study I, and values from the two studies are compared in Fig. 6. Repeated measures ANOVAs, with number of frequency bands and interferer condition as the within-subject variables and processing order as a between-subject variable, were conducted separately for binaural and monaural conditions. There was no significant effect of processing order on SRM. A main effect of number of frequency bands was found for the binaural $[F(3,16)=7.493,\ p<0.0001]$ but

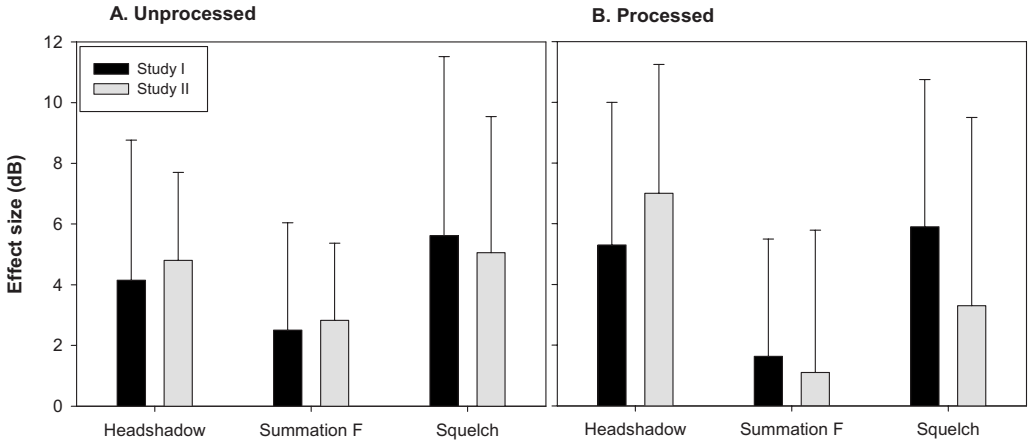


FIG. 7. Group means (+1 SD) are shown for head shadow, binaural summation in front, and binaural squelch. Results from Study I and Study II are compared for the unprocessed conditions (panel A) and the processed conditions (Panel B).

not for the monaural conditions. *Post-hoc t*-tests for this effect revealed that the amount of SRM for the unprocessed conditions was smaller than for the 16- ($p<0.01$), 8- ($p<0.0001$), and 4-band ($p<0.005$) conditions. Additionally, greater SRM was found for the 8-band than the 16-band condition ($p<0.0001$). SRM for the 4-band condition was comparable to SRM in the 16- and 8-band conditions.

A significant main effect of interferer condition was found for the monaural mode [$F(1,16)=150.412$, $p<0.0001$], where SRM for the left configuration was larger than that for the right configuration due to the presence of head shadow; there was no main effect of interferer condition for the binaural mode.

3. Bilateral effects

As in Study I, values for head shadow, binaural squelch, and binaural summation were calculated. A one-way repeated measures ANOVA was conducted on data from Study II for each effect, with the variable of interest being number of frequency bands (unprocessed, 16, 8, and 4). Similar to findings in Study I, none of the effects showed a dependence on number of frequency bands. However, to evaluate the effect of processing order on head shadow, summation, and squelch across the two studies, only data within the processed conditions were examined. Thus, data from each study within the processed conditions were pooled to yield an overall measure for each of the three effects. These pooled data are compared in Fig. 7 (panel B). A one-way ANOVA with processing order as the independent factor did not reveal statistically significant differences between the two studies. Differences in the squelch effect approached significance ($p=0.06$). The effect sizes for the unprocessed conditions (common across the two studies) are shown in Fig. 7 (panel A). The differences between the two subject groups were small and not statistically significant.

IV. DISCUSSION

A. Effect of CI vocoding on speech intelligibility

The current work shows that the unprocessed speech yielded the lowest average thresholds, in particular, in the conditions with interfering speech. This result is consistent with previous studies demonstrating that NHLs are less challenged by speech recognition in noise than subjects with hearing impairments (e.g., Chung and Mack, 1979; Dubno *et al.*, 1989; Pekkarinen *et al.*, 1990) or CI users (Nelson and Jin, 2004; Stickney *et al.*, 2004; Fu and Nogaki, 2005). This also confirms the well-known robustness of natural speech as a communication medium over spectrally degraded speech due to the listeners having access to spectral and temporal cues that have been shown to be important for speech recognition in noise (Assmann and Summerfield, 1990; Leek and Summers, 1993; Eisenberg *et al.*, 1995; Vliegen and Oxenham, 1999; Summers and Molis, 2004).

In the vocoder conditions speech recognition improved as the number of frequency bands was increased, which is consistent with previous reports using vocoded speech (Dorman and Loizou, 1997; Loizou *et al.*, 1999; Dorman *et al.*, 2000; Friesen *et al.*, 2001; Qin and Oxenham, 2003; Stickney *et al.*, 2004). In contrast with previous work (e.g., Dorman *et al.*, 1998), the present study demonstrated that 16 frequency bands are not sufficient to reach comparable performance to that with natural speech and that a larger number of bands is needed to support the dynamics of listening in complex auditory environments. In general, the elevated SRTs observed with a relatively large number of frequency bands underscore the importance of attempting to recapture and preserve information that is currently missing in today's clinical processors in order to enhance performance. Current CI processors encode temporal envelope cues to extract speech and discard fine-structure information, the latter perhaps being particularly important for listening to speech in the presence of interfering sounds (Rosen, 1992; Nelson *et al.*, 2003; Nie *et al.*, 2005; Fu and Nogaki, 2005).

B. Effect of CI vocoding on masking

Results showed that amount of masking increased considerably when spectrally degraded signals were used relative to the unprocessed conditions. However, amount of masking was comparable for the three frequency band conditions (16, 8, and 4) suggesting that once speech is spectrally degraded, the susceptibility to masking is relatively

high. In line with previous studies, current results underscore the limitation of CI vocoders in reproducing the fine structure information which is known to be important in speech recognition, particularly in the presence of temporally overlapping speech sounds (e.g., Smith *et al.*, 2002; Stickney *et al.*, 2005, 2007; Rubinstein and Hong, 2003; Wilson *et al.*, 2003, 2005). These results further suggest that when fine-structure information is reduced by vocoding, increasing the number of bands might not be the most constructive solution to the problem.

A further noteworthy finding is the relationship between masking and listening mode across the different spatial configurations. Overall, current results are consistent with reports in NHLs which showed that binaural hearing provides advantages over monaural hearing, in particular, in adverse listening environments (e.g., MacKeith and Coles, 1971; Bronkhorst and Plomp, 1988; Arsenault and Punch, 1999; Hawley *et al.*, 2004). Here we observed less masking under binaural than monaural conditions (see Fig. 2), with a dependence on the interferer location. When the interferer was placed in front, i.e., in absence of spatial separation between the target and interferer, binaural advantages were not observed.

C. Effect of CI vocoding on spatial release from masking

SRM increased as the number of frequency bands was reduced from unprocessed to 16- and then 8-bands. These results may be related to the finding in the normal-hearing literature that saliency of spatial cues increases as the listening environment becomes more difficult and complex. One such example is when the interferers carry linguistic content and context rather than consisting of noise or when the number of interfering talkers increases from 1 to 3 (e.g. Hawley *et al.*, 2004). Similarly, advantages of spatial cues become particularly robust when the target speech and interferers are comprised of identical or highly similar talkers, that is, when informational masking is present (e.g., Freyman *et al.*, 1999).

In the current studies, target-interferer similarity would have also been heightened when the number of frequency bands was reduced, leading to increase in informational masking. Thus, the increase in SRM in the vocoded conditions compared with the unprocessed conditions can be reasonably interpreted within the context of informational masking. What cannot be readily explained within that context is the fact that the increase in SRM was non-monotonic, declining for the 4-band conditions compared with the 8-band conditions. Perhaps, one interpretation is that the reduced SRM in the 4-band condition might have resulted from non-linearity in informational masking under conditions of degraded speech. As can be seen in Fig. 1, targets had to be presented at positive SNR values in order for listeners to achieve optimal performance, which reflects the difficulty of the task in that condition. It appears that when target recognition requires SNRs above 0 dB, listeners' susceptibility to informational masking is reduced, thereby minimizing SRM (Arbogast *et al.*, 2005; Freyman *et al.*, 2008).

The size of the bilateral effects measured with these simulations should be applied to actual CI users only with caution. The preservation and enhancement of SRM in the 8- and 16-band conditions was most likely due to availability of coordinated inputs across the two ears, such as identical frequency inputs at specific stimulation bands. As discussed below, simply smearing the spectral cues by processing HRTFs through CI simulation filters, as done in Study II, did not appear to have an effect on SRM. Thus, other aspects of signal processing in CIs need to be considered in order to understand how the gap between NHLs and CI users might be bridged.

D. Spatial effects related to the bilateral CI literature

Results from this study were analyzed in comparable ways to analyses that are typically conducted when bilateral CI users are tested. Three effects that are thought to reflect advantages stemming from having bilateral hearing were found to be significant: head shadow, binaural squelch, and binaural summation. The effects occurred regardless of the number of frequency bands in the signal, suggesting that benefits arising from bilateral hearing are not intimately dependent on frequency resolution.

In terms of benefits that are known to occur when two ears are activated, the head shadow is one of the largest. This effect occurs when the head of a listener acts as an acoustic barrier ("shadow") such that masker levels are attenuated at the ear contralateral to that of the masker location, improving SNRs at the contralateral ear. In a cocktail party environment, the benefit would arise when the target is near the "better" ear and the masker is contralateral to that ear (compared with ipsilateral to that ear). In the present study, when averaged across conditions, the head-shadow effect was approximately 5 dB, which is similar in magnitude to what has been found in persons with bilateral CIs (Gantz *et al.*, 2002; Muller *et al.*, 2002; Tyler *et al.*, 2002; Van Hoesel *et al.*, 2002; Schleich *et al.*, 2004; Litovsky *et al.*, 2006), and somewhat smaller than the 9–11 dB reported in NHLs (Arsenault and Punch, 1999; Bronkhorst and Plomp, 1988). The extent to which these differences are related to the choice of speech stimuli used in this study cannot be determined based on the current results; this topic should certainly be addressed in future studies. However, the similarity of the effect size between bilateral CI listeners and NHLs tested using binaural vocoded speech suggests that the vocoding reduced the effect size in a manner that effectively mimicked perceptual effects that arise when head-shadow cues are available to CI users.

Another effect measured in this study is the squelch effect, quantified for each subject as $[(\text{monaural})_{\text{left}}-(\text{binaural})_{\text{left}}]$. This effect, thought to be helpful for source segregation when sounds are spatially separated, requires that the auditory system make use of the differences in signals arriving at the two ears. In the present studies participants showed a squelch effect averaging approximately 6 dB in Study I and 3.6 dB in Study II, which is highly similar to the range of effect sizes (3–7 dB) reported in NHLs with undegraded stimuli (Levitt and Rabiner, 1967; Arsenault and Punch, 1999; Bronkhorst and Plomp, 1988; Hawley *et al.*,

2004). Our finding, that squelch occurs with either processing order, suggests that the smaller number of spectral bands available to CI users is not likely to be the limiting factor for eliciting the squelch effect.

If binaural temporal fine-structure cues are important for squelch, then one might have expected that the removal of those cues, as was done in Study II, would reduce or obliterate the effect. Statistical comparison of squelch between the two studies approached, but did not reach, significance, suggesting that the effect of processing order was weak or absent, or that variability in the data obscured the effect. However, the trend for squelch to be smaller in Study II, where the binaural cues in the stimuli were smeared by the vocoder, may help to understand the very small (Muller *et al.*, 2002; Schön *et al.*, 2002; Schleich *et al.*, 2004; Litovsky *et al.*, 2006) or absent (Gantz *et al.*, 2002; Van Hoesel and Tyler, 2003; Van Hoesel *et al.*, 2002) squelch seen in bilateral CI users. Clearly, lack of significant statistical effects here temper this conclusion and suggest that further work is needed in this area. Another, perhaps more likely, explanation for small squelch effects in bilateral CI users is the lack of coordinated inputs across the two ears and minimal or absent interaural timing cues. It has been reported that interaural level cues are the predominant cues for CI users (van Hoesel, 2004); this suggests that future advances in speech processors should include mechanism for restoring interaural timing cues.

In the third effect, known as binaural summation or redundancy, the signals reaching the two ears are very similar or identical, as the auditory stimulus is presented from 0° (front). The effect sizes in our studies were similar to those obtained in studies with hearing impaired listeners (Bronkhorst and Plomp, 1989) and were unaffected by order of processing. The summation effect is also an effect that is found in some but not all bilateral CI users (Schleich *et al.*, 2004; Litovsky *et al.*, 2006). In the study of Litovsky *et al.* (2006) 15/34 subjects (44%) demonstrated this effect when compared with either of the two ears alone, while 17/34 (50%) subjects had no effect, and 2/34 subjects (6%) showed a decrement in the bilateral condition rather than improved performance. Thus, like squelch, the summation effect might be a good example of a benefit that comes from having inputs present at both ears and depends on highly symmetrical (or identical) hearing integrity, and possibly also coordinated timing of inputs between the ears. These are factors that are known to be problematic in bilateral CI users, but that were clearly surmounted by the stimulation approaches used here.

E. Effect of simulation order on utility of spatial information

By varying the order of stimulus processing in the two studies, we were able to examine effects of two aspects of CI simulation; removal of speech fine-structure cues (Study I) and subsequently also removal of binaural temporal fine-structure cues (Study II). Because the stimulation to the two ears was coordinated, as occurs in NHLs, interaural timing and level differences in the envelope remained unperturbed, which renders these good candidates for cues used by the listeners. A main effect of processing order was not found for SRTs. However, a significant interaction of number of frequency bands by processing order revealed that for the 8-band conditions binaural SRTs obtained from Study I were lower than those obtained in Study II (in which HRTFs were processed through the CI vocoder). In the 16-band condition, i.e., the condition with substantially richer spectral information, there was no effect of processing order. Together with the lack of interaction effects in the monaural conditions, these results suggest that under binaural conditions, spectral information that is available in the HRTFs might be useful for speech recognition in adverse listening environments. The underlying mechanisms responsible for this are likely to be ones that produce redundancy or summation of information that is required for speech perception.

The consequence of higher SNRs being required for listeners to achieve optimal performance is consistent with what is known about CI users and the challenges that they face in noisy situations. This finding has implications for true CI users who are shown to utilize up to eight independent channels of spectral information (Friesen *et al.*, 2001). Additionally, it is important to note that differences in SRTs across the two studies were not observed for the 4-band conditions. This is most likely due to the severe degradation of spectral information available in the 4-band conditions regardless of the presence of HRTF cues. On the other hand, the extent to which inter-subject variability precluded these differences cannot be ruled out. The lack of statistically significant effects in the monaural conditions suggests that directionally dependent cues in the HRTFs, which were available in Study I but were most likely eliminated in Study II by processing HRTFs through the vocoder, may not have served an important purpose for the effects studied here.

Of further interest is whether the amount of SRM is dependent on the preservation of directional cues that are available in the HRTFs. SRM was not significantly different across the two processing approaches. In Fig. 6, however, one can see a trend for smaller SRM in the right-masker configuration in Study II than in Study I. It is noteworthy that the data have an inherent level of inter-subject variability. Individual differences are a hallmark of some noted perceptual phenomena such as informational masking (e.g., Oh and Lutfi, 1998; Durlach *et al.*, 2003), which, as discussed above, seems to have arisen in the vocoded stimuli used in the current experiments. The choice of using a speech interferer in these experiments was based on the desire to utilize a stimulus paradigm that more realistically represents real-world listening situations encountered by CI users in everyday situations. The extent to which the inter-subject variability might have been so large as to obscure effects due to signal processing order or other manipulations conducted here might be investigated in future studies perhaps with fewer conditions but a much larger *N* size for participants. Alternatively, one might tackle this issue by using stimuli that are constructed so as to maximize energetic, rather than informational, masking.

Regardless of the variability, the spatial effects observed here were either comparable to or greater than those reported

in bilateral CI users. These results suggest that directional cues that exist at the output of the vocoder, even after the HRTFs have been processed, are sufficient for the occurrence of spatially dependent bilateral benefits. The present study demonstrated that, by preserving binaurally coordinated stimulation in the envelopes of the signals alone, benefits from bilateral CIs could be substantial, regardless of the amount of spectral degradation in the speech signal. This suggests that, while preservation of fine-structure in the signal may offer other benefits, envelope-based binaural differences are likely to offer a substantial portion of the advantage for listening in complex environments. The extent to which fine-structure vs envelope cues might each contribute to improved performance is obviously an important topic for further investigation.

V. SUMMARY AND CONCLUSION

The current study examined the effect of spectral resolution on speech intelligibility and SRM in binaural and monaural listening conditions in NHLs. The order of signal processing of the vocoded speech and the directionally dependent HRTFs had little effect on the results. The findings are consistent with the notion that increased spectral information is important for improved speech intelligibility. However, the benefit of spatial cues was most pronounced under conditions of spectral degradation of speech, when the target and interfering speech are more likely to be confused and thus when informational masking is likely to be larger. Benefits from binaural hearing that are rarely observed in true bilateral CI users were seen here. This suggests that for the effects studied in these experiments, preservation of binaural coordination between the two ears may be important to support bilateral implantation.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Richard Freyman and two anonymous reviewers for their helpful comments on earlier drafts of this manuscript. The authors are grateful to Shelly Godar and Tanya Jensen for assisting with subject recruitment and to Lindsey Rentmeester and Nick Liimatta for assisting with data collection. They would also like to thank Christopher Long for his feedback on an earlier version of the manuscript. Portions of this work were presented at the 2006 Meeting of the Association for Research in Otolaryngology. Work supported by NIH-NIDCD Grant No. R01DC030083 to R.Y.L.

[1]HRTFs from NHLs such as those used here are typically measured in the ear canal (Blauert, 1997), and contain high-frequency cues that are not preserved by the CI processors.

[2]In typical cochlear implant systems, the highest frequency is approximately 8000 which is lower than the 10,300 Hz cutoff used here. However, logarithmically this value is not that much higher than the values used in current CIs. There is good reason for providing higher frequencies because localization cues that result from directionally dependent filtering of sounds by the head and ears are greatest at frequencies between the range of 8–10 000 Hz than at the lower frequencies. Some of the MAPS that are provided by manufacturers do offer this range as an option (e.g., Table 9 in the cochlear system). Finally, the higher frequency cutoff used here was selected for consistency and comparability with results from other studies being conducted by our group in which sound localization ability is investigated using the same stimuli.

Arbogast, T., Mason, C., and Kidd, G., Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Arbogast, T., Mason, C., and Kidd, G., Jr. (**2005**). "The effect of spatial separation on informational masking of speech in normal hearing and hearing impaired listeners," J. Acoust. Soc. Am. **117**, 2169–2180.

Arsenault, M., and Punch, J. (**1999**). "Nonsense-syllable recognition in noise using monaural and binaural listening strategies," J. Acoust. Soc. Am. **105**, 1821–1830.

Assmann, P., and Summerfield, Q. (**1990**). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.

Assmann, P., and Summerfield, Q. (**1994**). "The contribution of waveform interactions to the perception of concurrent vowels," J. Acoust. Soc. Am. **95**, 471–484.

Bacon, S., Opie, J., and Montoya, D. (**1998**). "The effect of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds," J. Speech Lang. Hear. Res. **41**, 549–563.

Başkent, D., and Shannon, R. V. (**2007**). "Combined effects of frequency compression-expansion and shift on speech recognition," Ear Hear. **28**, 277–289.

Battmer, R., Feldmeier, I., Kohlenberg, A., and Lenarz, T. (**1997**). "Performance of the new Clarion speech processor 1.2 in quiet and in noise," Am. J. Otol. **18**, S144–S146.

Bird, J., and Darwin, C. J. (**1998**). "Effects of a difference in fundamental frequency in separating two sentences," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London).

Blauert, J. (**1997**). *Spatial Hearing—Revised Edition: The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA).

Bronkhorst, A., and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J. Acoust. Soc. Am. **83**, 1508–1516.

Bronkhorst, A., and Plomp, R. (**1989**). "Binaural speech intelligibility in noise for hearing-impaired listeners," J. Acoust. Soc. Am. **86**, 1374–1383.

Brungart, D. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1119.

Buss, S., and Florentine, M. (**1985**). "Gap detection in normal and impaired listeners: The effect of level and frequency," in *Time Resolution in Auditory Systems*, edited by A. Michelsen (Springer-Verlag, London), pp. 159–179.

Chung, D., and Mack, B. (**1979**). "The effect of masking by noise on word discrimination scores in listeners with normal hearing and with noise-induced hearing loss," Scand. Audiol. **8**, 139–143.

Dorman, M., Loizou, P., Fitzke, J., and Tu, Z. (**1998**). "The recognition of sentences in noise by normal hearing listeners using simulations of cochlear implant signal processors with 6–20 channels," J. Acoust. Soc. Am. **104**, 3583–3585.

Dorman, M., and Loizou, P. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for normal hearing listeners and patients with cochlear implants," Am. J. Otol. **18**, S113–S114.

Dorman, M., Loizou, P., Kemp, L., and Kirk, K. (**2000**). "Word recognition by children listening to speech processed into small number of channels, data from normal-hearing children and children with cochlear implants," Ear Hear. **21**, 590–596.

Drullman, R., and Bronkhorst, A. (**2000**). "Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation," J. Acoust. Soc. Am. **107**, 2224–2235.

Dubno, J., Dirks, D., and Schaefer, A. (**1989**). "Stop-consonant recognition for normal hearing listeners and listeners with high frequency loss II: Articulation index predictions," J. Acoust. Soc. Am. **85**, 355–364.

Durlach, N., Mason, C., Shinn-Cunningham, B., Arbogast, T., Colburn, H., and Kidd, G., Jr. (**2003**). "Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," J. Acoust. Soc. Am. **114**, 368–379.

Eisenberg, L., Dirks, D., and Bell, T. (**1995**). "Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing," J. Speech Hear. Res. **38**, 222–233.

Freyman, R., Balakrishnan, U., and Helfer, K. (**2008**). "Spatial release from

masking with noise-vocoded speech," J. Acoust. Soc. Am. **124**, 1627–1637.
Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.
Friesen, L., Shannon, R., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implant," J. Acoust. Soc. Am. **110**, 1150–1163.
Fu, Q., and Nogaki, G. (**2005**). "Noise susceptibility of cochlear implant users, the role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**, 19–27.
Gantz, B., Tyler, R., Knutson, J., Woodworth, G., Abbas, P., McCabe, B., Hinrichs, J., Tye-Murray, N., Lansing, C., Kuk, F., and Brown, C. (**1988**). "Evaluation of five different cochlear implant designs: Audiologic assessment and predictions of performance," Laryngoscope **98**, 1100–1106.
Gantz, B., Tyler, R., Rubinstein, J., Wolaver, A., Lowder, M., Abbas, P., Brwon, C., Hughes, M., and Preece, J. (**2002**). "Binaural cochlear implants placed during the same operation," Otol. Neurotol. **23**, 169–180.
Gardner, W., and Martin, K. (**1994**). "HRTF measurements of a KEMAR dummy-head microphone," The MIT Media Laboratory Machine Listening Group, http://sound.media.mit.edu/resources/KEMAR.html (Last viewed February 2009).
Hawley, M., Litovsky, R., and Colburn, H. (**1999**). "Speech intelligibility and localization in a multi-source environment," J. Acoust. Soc. Am. **105**, 3436–3448.
Hawley, M., Litovsky, R., and Culling, J. (**2004**). "The benefits of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am. **115**, 833–843.
Iwaki, T., Matsushiro, N., Mah, S., Sato, T., Yasuoka, E., Yamamoto, K., and Kubo, T. (**2004**). "Comparison of speech perception between monaural and binaural hearing in cochlear implant patients," Acta Oto-Laryngol. **124**, 358–362.
Kidd, G., Jr., Mason, C., Rohtla, T., and Deliwala, P. (**1998**). "Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns," J. Acoust. Soc. Am. **104**, 422–431.
Kidd, G., Jr., Mason, C., and Arbogast, T. (**2002**). "Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns," J. Acoust. Soc. Am. **111**, 1367–1376.
Leek, M., and Summers, V. (**1993**). "The effect of temporal waveform shape on spectral discrimination by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **94**, 2074–2082.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Levitt, H., and Rabiner, L. (**1967**). "Binaural release from masking for speech and gain in intelligibility," J. Acoust. Soc. Am. **42**, 601–608.
Litovsky, R. Y. (**2005**). "Speech intelligibility and spatial release from masking in young children," J. Acoust. Soc. Am. **117**, 3091–3099.
Litovsky, R. Y., Parkinson, A., and Arcaroli, J. (**2009**). "Spatial hearing and speech intelligibility in bilateral cochlear implant users," Ear Hear. **30**, 419–431.
Litovsky, R. Y., Parkinson, A., Arcaroli, J., and Sammath, C. (**2006**). "Clinical study of simultaneous bilateral cochlear implantation in adults: A multicenter study," Ear Hear. **27**, 714–731.
Litovsky, R., Parkinson, A., Arcoroli, J., Peters, R., Lake, J., Johnstone, P., and Yu, G. (**2004**). "Bilateral cochlear implants in adults and children," Curr. Opin. Otolaryngol. Head Neck Surg. **130**, 648–655.
Loizou, P., Dorman, M., and Tu, Z. (**1999**). "On the number of channels needed to understand speech," J. Acoust. Soc. Am. **106**, 2097–3003.
Loizou, P., Hu, Y., Litovsky, R., Yu, G., Peters, R., Lake, J., and Roland, P. (**2009**). "Speech recognition by bilateral cochlear implant users in a cocktail-party setting," J. Acoust. Soc. Am. **125**, 372–383.
MacKeith, N., and Coles, R. (**1971**). "Binaural advantages in hearing of speech," J. Laryngol. Otol. **85**, 213–232.
Muller, J., Schon, F., and Helms, J. (**2002**). "Speech understanding in quiet and noise in bilateral users of the MED-EL COMBI 40/40+ cochlear implant system," Ear Hear. **23**, 198–206.
Muller-Deile, J., Schmidt, B., and Rudert, H. (**1995**). "Effects of noise on speech discrimination in cochlear implant patients," Ann. Otol. Rhinol. Laryngol. Suppl. **166**, 303–306.
Neff, D. (**1995**). "Signal properties that reduce masking by simultaneous, random-frequency maskers," J. Acoust. Soc. Am. **98**, 1909–1920.
Nelson, P., and Jin, S. (**2004**). "Factors affecting speech understanding in gated interference: Cochlear implant users and normal hearing listeners," J. Acoust. Soc. Am. **115**, 2286–2294.
Nelson, P., Jin, S., Carney, A., and Nelson, D. (**2003**). "Understanding speech in modulated interference: Cochlear implant users and normal hearing listeners," J. Acoust. Soc. Am. **113**, 961–968.
Nie, K., Stickney, G., and Zeng, F. G. (**2005**). "Encoding frequency modulation to improve cochlear implant performance in noise," IEEE Trans. Biomed. Eng. **52**, 64–73.
Oh, E., and Lutfi, R. (**1998**). "Nonmonotonicity of informational masking," J. Acoust. Soc. Am. **104**, 3489–3899.
Pekkarinen, E., Salmivalli, A., and Suonpaa, J. (**1990**). "Effect of noise on word discrimination by subjects with impaired hearing, compared with those with normal hearing," Scand. Audiol. **19**, 31–36.
Qin, M., and Oxenham, A. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.
Rosen, S. (**1992**). "Temporal information in speech acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.
Rothauser, E., Chapman, W., Guttman, N., Nordby, K., Silbigert, H., Urbanek, G., and Weinstock, M. (**1969**). "IEEE Recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 227–246.
Rubinstein, J., and Hong, R. (**2003**). "Signal coding in cochlear implants: Exploiting stochastic effects of electrical stimulation," Ann. Otol. Rhinol. Laryngol. Suppl. **191**, 14–19.
Schleich, P., Nopp, P., and D'Haese, P. (**2004**). "Head shadow, squelch, and summation effects in bilateral users of the Med-El COMBI 40/40+ Cochlear implant," Ear Hear. **25**, 197–204.
Schon, F., Muller, J., and Helms, J. (**2002**). "Speech reception thresholds obtained in a symmetrical four-loudspeaker arrangement from bilateral users of MED-EL cochlear implant," Otol. Neurotol. **3**, 710–714.
Shannon, R. V., Galvin, J. J. III, and Baskent, D. (**2002**). "Holes in hearing," J. Assoc. Res. Otolaryngol. **3**, 185–199.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Skinner, M., Clark, G., Whitford, L., Seligman, P., Staller, S., Shipp, D., Shallop, J., Everingham, C., Menapace, C., and Arndt, P. (**1994**). "Evaluation of a new spectral peak coding strategy for the Nucleus 22 channel cochlear implant system," Am. J. Otol. **15**, 15–27.
Smith, Z., Delgutte, B., and Oxenham, J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature (London) **416**, 87–90.
Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.
Stickney, G., Assmann, P., Chang, J., and Zeng, F. G. (**2007**). "Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences," J. Acoust. Soc. Am. **122**, 1069–1078.
Stickney, G., Nie, K., and Zeng, F. G. (**2005**). "Contribution of frequency modulation to speech recognition in noise," J. Acoust. Soc. Am. **118**, 2412–2420.
Summers, V., and Molis, M. (**2004**). "Speech recognition in fluctuating and continuous maskers: Effects of hearing loss and presentation level," J. Speech Lang. Hear. Res. **47**, 245–256.
Tyler, R., Dunn, C., Witt, S., and Preece, J. (**2003**). "Update on bilateral cochlear implantation," Curr. Opin. Otolaryngol. Head Neck Surg. **11**, 388–393.
Tyler, R., Summerfield, Q., Wood, E., and Fernandes, M. (**1982**). "Psychoacoustic and phonetic temporal processing in normal and hearing impaired," J. Acoust. Soc. Am. **72**, 740–752.
Tyler, R., Gantz, B., Rubinstein, J., Wilson, B., Parkinson, A., Wolaver, A., Precce, J., Witt, S., and Lowder, M. (**2002**). "Three-month results with bilateral cochlear implants," Ear Hear. **23**, 80S–89S.
Van Hoesel, R. (**2004**). "Exploring the benefits of bilateral cochlear implants," J. Nurs. Meas **9**, 234–246.
Van Hoesel, R., and Tyler, R. (**2003**). "Speech perception, localization, and lateralization with bilateral cochlear implants," J. Acoust. Soc. Am. **113**, 1617–1630.
Van Hoesel, R., Ramsden, R., and O'Driscoll, M. (**2002**). "Sound-direction identification, interaural time delay discrimination and speech intelligibility advantages in noise for a bilateral cochlear implant users," Ear Hear. **23**, 137–149.
Vliegen, J., and Oxenham, A. (**1999**). "Sequential stream segregation in the absence of spectral cues," J. Acoust. Soc. Am. **105**, 339–346.

Waltzman, S. B., Cohen, N. L., and Fisher, S. (**1992**). "An experimental comparison of cochlear implant systems," Semin. Hear. **13**, 195–207.
Wichmann, F. A., and Hill, J. (**2001a**). "The psychometric function: I. Fitting, sampling, and goodness of fit," Percept. Psychophys. **63**, 1290–1313.
Wichmann, F. A., and Hill, J. (**2001b**). "The psychometric function: II. Bootstrap-based confidence intervals and sampling," Percept. Psychophys. **63**, 1314–1329.
Wilson, B., Lawson, D., and Muller, J. (**2003**). "Cochlear implant: Some likely next steps," Annu. Rev. Biomed. Eng. **5**, 207–249.
Wilson, B., Schatzer, R., Lopez-Poveda, E., Sun, X., Lawson, D., and Wolford, R. (**2005**). "Two new directions in speech processor design for cochlear implants," Ear Hear. **26**, 73S–81S.

Relative influence of interaural time and intensity differences on lateralization is modulated by attention to one or the other cue: 500-Hz sine tones

Albert-Georg Lang[a] and Axel Buchner
Institut für Experimentelle Psychologie, Heinrich-Heine-Universität, 40225 Düsseldorf, Germany

(Received 9 January 2009; revised 30 July 2009; accepted 31 July 2009)

When interaural time differences and interaural intensity differences are set into opposition, the measured trading ratio depends on which cue is adjusted by the listener. In an earlier article [Lang, A.-G., and Buchner, A., J. Acoust. Soc. Am. **124**, 3120–3131 (2008)], four experiments showed that the perceived localization of a broad band sound for which differences in one cue were compensated by differences in the other cue such that the sound seemed to originate from a central position shifted back toward the location from which the sound appeared to originate before the adjustment. It was argued that attention shifted toward the effect of the to-be-adjusted cue during the compensation task, leading to an increased weighting of the to-be-adjusted cue. The use of broadband stimuli raises the question whether the "shift-back effect" was caused by attentional shifts to the effect of the to-be-adjusted binaural cue or by attention shifts to the particular frequency range which is most important for localizations based on the to-be-adjusted cue. Two experiments are reported in which sine tones of 500 Hz were used instead of broadband sounds. The shift-back effect could still be observed, supporting our original hypothesis. A control experiment showed that participants had accurate representations of the critical central position.
 [DOI: 10.1121/1.3212927]

PACS number(s): 43.66.Qp, 43.66.Ba, 43.66.Pn [JCM] Pages: 2536–2542

I. INTRODUCTION

In an earlier article (Lang and Buchner, 2008), we reported four experiments showing that the equivalence relation[1] between interaural time differences (ITDs) and interaural intensity differences (IIDs) varies as a function of the task participants have to perform. The experiments consisted of two phases. In the *compensation phase*, participants canceled out the effect of one preset binaural cue on localization by adjusting a compensatory value of the other cue until the sound was located in a central position. In the *localization phase*, participants assessed the virtual position of the sound, using the preset value of the fixed cue and using the same value of the complementary cue as previously adjusted. The sounds were no longer perceived as originating from the center. Instead, their perceived location was shifted back toward the location from which they appeared to originate before the adjustment. This "shift-back effect" suggests that the to-be-adjusted cue received a larger weight than the other cue during the compensation task.

Specifically, while adjusting one binaural cue in order to compensate for the effect of the other cue participants moved a control element and simultaneously received feedback about the effects of their adjustments in terms of immediate changes in the virtual location of the sound source. Given that participants were instructed to find an adjustment value that led to a certain localization (at the central position), they had to carefully observe the relation between their adjustments and the perceived changes in the sound source location. We presumed that this led to an increased attention to the effect of participants' adjustments on perceived location which, in turn, led to an increased perceptual weight of the adjusted cue in relation to the complementary cue. Let us take a closer look at the mechanism that we assume to be responsible for the effect. Attention—in terms of a resource that can be used strategically—can be directed only to stimulus features that participants can distinguish and that they could report if they were asked to. Attention, in this sense, was directed toward the effects of participants' adjustments of the control element on the virtual location of the sound source. This focusing of attention is assumed to have led to an increased weighting of the binaural cue associated with the control element in the process of ITD and IID information integration. Thus, directed attention led to a shift in the relative weighting of the binaural cues which then automatically affected sound source location. Note that there is some evidence in the literature that setting both binaural cues into opposition may lead to the occurrence of two images, a time image and a time/intensity image (e.g., see Whitworth and Jeffress, 1961; Hafter and Jeffress, 1968). A possible variant of our attentional explanation of the shift-back effect would thus be that participants attend to the image whose position was effected by their adjustments.

A plausible alternative explanation of the shift-back effect starts by noting that we (Lang and Buchner, 2008) previously used wideband stimuli (a female voice) even though in most studies 500-Hz sine tones were used (for an overview, see Trahiotis and Kappauf, 1978). By using broadband stimuli, we expected more precise localization judgments (e.g., see Stevens and Newman, 1936) and, hence, an increased chance of finding a possible shift-back effect. Nevertheless, the trading ratios found in our experiments were

[a]Author to whom correspondence should be addressed. Electronic mail: albert.lang@uni-duesseldorf.de

quite similar to those of the trading experiments reported by Young and Levine (1977) where 500-Hz sine tones were used. When ITD was the preset cue that was to be compensated by a complementary IID, we found a trading ratio of 80.1 μs/dB for a preset ITD of 600 μs which compares nicely with a ratio of 79.4 μs/dB for a preset ITD of 500 μs in Young and Levine (1977). When the roles of ITD and IID were interchanged, we found a trading ratio of 27.7 μs/dB for a preset IID of 7.5 dB which again seems to fit with a ratio of 40.4 μs/dB for a preset IID of 8 dB in Young and Levine (1977).

According to Rayleigh's "Duplex Theory" of sound localization (Strutt, 1907) IIDs are the more important interaural cue for sound localization of high-frequency sounds while ITDs are more important for low-frequency sounds. In spite of the considerable age of Rayleigh's theory, it is still in good agreement with actual findings (e.g., see Macpherson and Middlebrooks, 2002). This "specificity" of the binaural cues to high- or low-frequency ranges poses an alternative explanation of the shift-back effect in the experiments reported by Lang and Buchner (2008). It may be hypothesized that the compensation task did not lead to a shift of attention to the effect of one of the two binaural cues on localization but to a shift of attention to one of two *frequency ranges*. ITDs can only be evaluated at lower frequencies because of a loss of phase-locking in the auditory nerve at high frequencies (Macpherson and Middlebrooks, 2002). In natural hearing situations, IIDs mainly occur with high frequencies since low frequencies become deflected around the listener's head. Thus, life-long learning experience of IID-based sound localization could have lead to a stronger association between high frequencies and IIDs than between low frequencies and IIDs. This stronger association may influence localization judgments even in a situation where low-frequency IIDs are available (as is the case in our headphone-based experiments). In our experiments, listeners had to adjust the IID of a broadband sound in order to compensate for an ITD. To do so they moved a control element and simultaneously received feedback in terms of a change in the virtual sound source location. Present during decades of sound localization experience, the ubiquitous association between IIDs and high-frequency sound components could have led to a shift of attention toward the high-frequency components of the sounds in our experiments in order to receive the best feedback about the relation between their adjustments and the changes in the sound source position. Similarly, while adjusting the ITD in order to compensate for an IID listeners' attention could have shifted toward the low-frequency components of the sounds.

If this assumption were correct, then the shift-back effect found in the experiments reported by Lang and Buchner (2008) should no longer be observed if pure tones are used instead of broadband stimuli. However, if our original hypothesis were correct that shifts of attention between the binaural cues *themselves* (more precisely: the effect of either cue on localization) caused the shift-back effect, then the shift-back effect should also emerge when pure tones are used. Two experiments were conducted to test these predictions of the two alternative explanations of the shift-back effect. In experiment 1, participants compensated preset ITDs by complementary IIDs. In experiment 2, the roles of ITDs and IIDs were interchanged.

II. EXPERIMENT 1

A. Method

1. Participants

Participants were 12 female and 6 male persons, most of whom were students at Heinrich-Heine-Universität Düsseldorf. Their age ranged from 19 to 42 years ($M=26.0$, $SD=5.9$). All participants reported normal hearing. They were paid for participating or received course credit.

2. Apparatus, stimuli, and procedure

The experiment was a replication of experiment 1a of Lang and Buchner (2008) with the only difference being that 500-Hz sine tones were used as stimuli instead of natural speech sounds. In order to maximize the precision with which ITDs could be regulated the tones were sampled at a resolution of 32 bits at 96 kHz. During the experiment, the sounds were presented via headphones (AKG K-501) at a sound level of about 60 dB_{SPL} (*A*-weighted).

The experiment consisted of two phases, a compensation phase in which participants compensated a preset ITD by an IID of inverse sign and a localization phase that consisted of pure localization judgments. Each trial of the compensation phase started with a continuous loop in which the sine tone was presented for 1000 ms alternating with 1250 ms of silence. In order to avoid steep transients squared cosine ramps of 50 ms rise and fall time were used at the beginning and at the end of the tones. The tones were presented with one of seven preset ITDs (−600, −400, −200, 0, 200, 400, or 600 μs)[2]; each preset ITD was presented in five trials, such that there were $7\times5=35$ compensation trials. The control element that was used by participants to choose a compensatory IID was a vertical slider displayed on a computer monitor which controlled the level difference between the left and right headphone within a range of ±15 dB. Each trial began at a randomly chosen starting position of the slider. When participants had finished the adjustment they clicked on a "Continue" button in order to start the next trial.

The *localization phase* consisted of 35 critical and 35 control trials. Each trial was presented with a preset ITD (−600, −400, −200, 0, 200, 400, or 600 μs). The IID was set to 0 dB in all control trials; in the critical trials the IID was identical with the IID, the participant had chosen during the parallel trial of the compensation phase. On the computer monitor, a sketch of a human head wearing headphones was displayed as seen from behind such that the left side of the sketch paralleled the left side of the participant's head. A red dot could be moved to angles between −90° and +90° on the upper hemicycle of the displayed head using the computer mouse.

3. Design

The independent variable was the ITD, which was manipulated within-subject in seven steps (−600, −400, −200, 0,

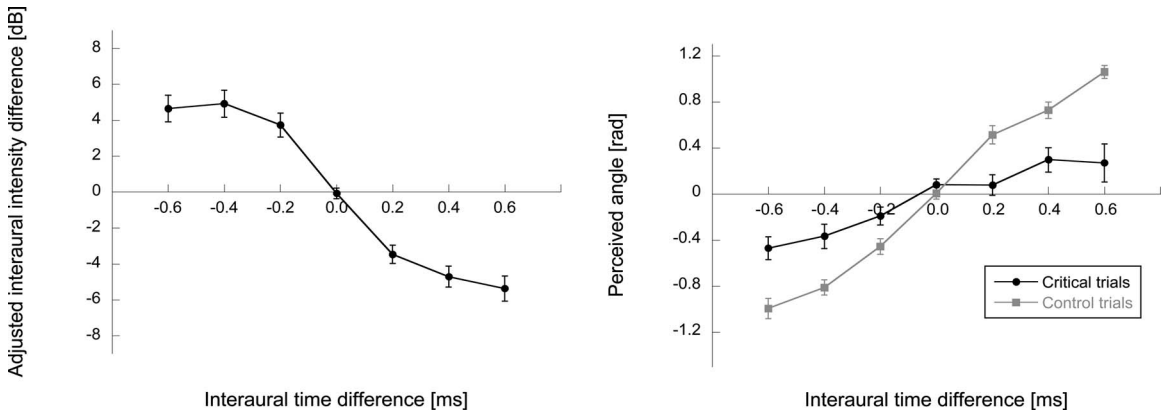


FIG. 1. Left panel: IIDs chosen to compensate for preset ITDs during the compensation phase of experiment 1 (error bars denote standard errors of the means). Right panel: Relation between the preset ITD and the perceived location during the critical trials and the control trials of the localization phase of experiment 1 (error bars denote standard errors of the means).

200, 400, and 600 μs). Dependent variables were the IID chosen during the compensation phase and the perceived location during the localization phase. A multivariate analysis of variance approach (MANOVA) was used for within-participant comparisons. Polynomial contrasts were evaluated from order 1 to order 4. Partial η^2 is reported as an effect size measure.

B. Results

Figure 1 (left panel) illustrates the relation between the preset ITD and the IID chosen to compensate for the effect of the ITD during the compensation phase. A MANOVA showed that the effect of the preset ITD (−600, −400, −200, 0, 200, 400, and 600 μs) on the chosen IID was statistically significant [$F(6,12)=7.89$, $p=0.001$, $\eta^2=0.80$]. An analysis of the polynomial contrasts revealed statistically significant first and third order trends [$F(1,17)=60.04$, $p<0.001$, $\eta^2=0.78$; $F(1,17)=35.08$, $p<0.001$, $\eta^2=0.67$, respectively].

Figure 1 (right panel) shows the relation between the preset ITD and the perceived location during the localization phase. A MANOVA for the control trials showed a significant effect of the preset ITD on perceived sound source location [$F(6,12)=74.0$, $p<0.001$, $\eta^2=0.97$].

A more interesting analysis concerns the perceived sound source locations of the critical trials. If there were no shift-back effect, then all localization judgments of the critical trials should be at zero; that is, the graph of the critical trials should lie on the abscissa. However, the fact that the slope of the graph of the critical trials is positive indicates that localization judgments were dependent on the preset ITD and hence, that the shift-back effect is present. A MANOVA for the critical trials showed that the effect of the ITD was indeed significant [$F(6,12)=3.62$, $p=0.028$, $\eta^2=0.64$]. An analysis of the polynomial contrasts showed that only the linear component was statistically significant [$F(1,17)=18.21$, $p=0.001$, $\eta^2=0.52$].

C. Discussion

The central result of experiment 1 is that our earlier findings could be replicated: The shift-back effect demonstrated in experiment 1a of Lang and Buchner (2008) with natural speech sounds also emerged in the current experiment with 500-Hz sine tones.

As a side note, in the critical trials of the localization phase of the current experiment 1, the effect of the preset ITD on perceived localization angle ($\eta^2=0.64$) was somewhat smaller than the effect in experiment 1a of Lang and Buchner (2008) ($\eta^2=0.85$). The relevant data are displayed in the right panel of Fig. 2. Of course, this may just represent random variation between experiments with different samples of participants so that we cannot be sure that the difference observed here warrants a substantive interpretation. That being said, the left panel of Fig. 2 shows another difference between these experiments in terms of the IIDs chosen by participants to compensate for preset ITDs during the compensation phases. In the current experiment 1, the chosen IIDs were smaller than in experiment 1a of Lang and Buchner (2008). It appears as if smaller IIDs were perceived as being sufficient to compensate for the effects of preset ITDs when 500-Hz sine tones were used rather than natural speech sounds. This may imply that in the compensation phase of the current experiment 1, ITDs received a smaller perceptual weight than in experiment 1a of Lang and Buchner (2008), or alternatively, that IIDs received a larger perceptual weight in the current experiment, or both.

III. EXPERIMENT 2

A. Method

1. Participants

Participants were 43 female and 4 male persons, most of whom were students at the Heinrich-Heine-Universität Düsseldorf. Their age ranged from 18 to 48 years ($M=24.4$, $SD=7.3$). All participants reported normal hearing. They were paid for participating or received course credit.

2. Apparatus, stimuli, and procedure

Experiment 2 was a replication of experiment 2 in Lang and Buchner (2008) with 500-Hz sine tones as stimuli and is identical to experiment 1 except for the fact that the roles of ITDs and IIDs were interchanged. Each trial was presented with a preset IID of −7.5, −5.0, −2.5, 0, 2.5, 5.0, or 7.5 dB.

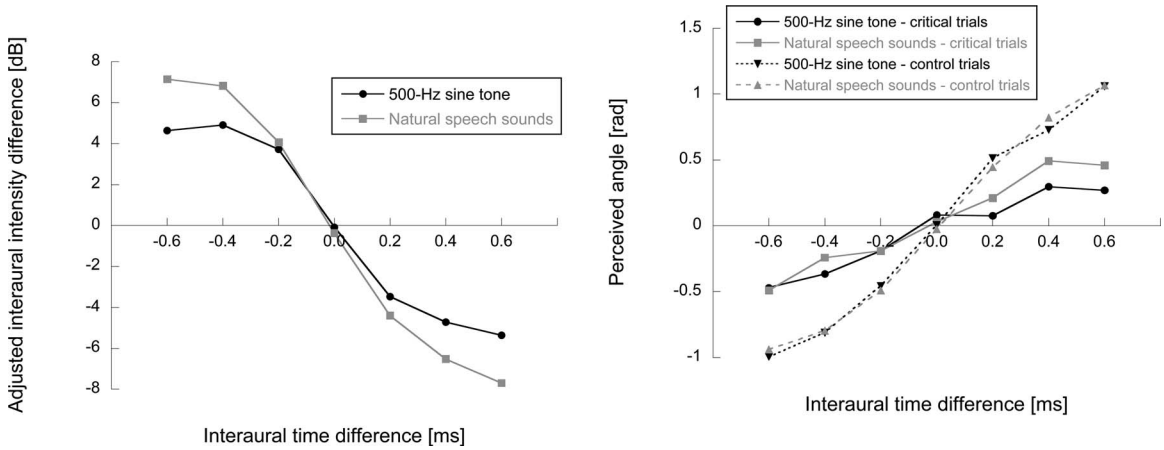


FIG. 2. Left panel: IIDs chosen to compensate for given ITDs during the compensation phase of experiment 1 (500-Hz sine tone) and the compensation phase of experiment 1a of Lang and Buchner (2008)(natural speech sounds). Right panel: Relation between the preset ITD and perceived location during the critical trials and the control trials of the localization phases of both experiments.

The slider that was used during the compensation trials allowed adjustments of the ITD between −600 and +600 μs. As in experiment 1, a sampling resolution of 32 bits/96 kHz was used.

During a pilot study prior to experiment 2 in Lang and Buchner (2008), the subjective impression was noted that for some trials even an ITD of ±600 μs would not seem to compensate the given IID. For this reason, a checkbox labeled "Not enough" was displayed next to the slider, just as in experiment 2 of Lang and Buchner (2008). Participants were instructed to check the box if they had the impression that even the most extreme slider position was not sufficient to achieve a sound localization on the midline. These trials were excluded from all further analyses since the occurrence of a shift-back effect would have been a trivial finding in these trials.

3. Design

The independent variable was the preset IID, which was manipulated within-subject in seven steps (−7.5, −5.0, −2.5, 0, 2.5, 5.0, and 7.5 dB). Dependent variables were (a) the ITD chosen during the compensation phase and (b) the perceived location during the localization phase of the experiment.

B. Results

Eleven participants were excluded because for one or more preset IIDs, they chose the "Not enough" checkbox in all trials (i.e., none of the five trials with a specific preset IID could be compensated for localization in the center). For the remaining participants in 7.7% of all trials the "Not enough" checkbox was chosen. "Not enough" was chosen most frequently when the preset IID was ±7.5 dB. Of all "Not enough" trials of all participants (including the 11 excluded participants) 73.8% occurred with ±7.5 dB; 20.1%, 3.5%, 2.5%, and 1.99% of the "Not enough" choices were associated with a preset IID of ±5, ±2.5, and 0 dB, respectively.

The left panel of Fig. 3 illustrates the relation between the preset IID and the ITD chosen to compensate for the effect of the IID during the compensation phase. A MANOVA showed that the effect of the preset IID (−7.5, −5.0, −2.5, 0, 2.5, 5.0, and 7.5 dB) on the chosen ITD was statistically significant [$F(6,30)=17.1$, $p<0.001$, $\eta^2=0.77$]. An analysis of the polynomial contrasts showed that the linear and the cubic components were statistically significant [$F(1,35)=63.32$, $p<0.001$, $\eta^2=0.64$ and $F(1,35)=6.35$, $p=0.016$, $\eta^2=0.15$].

Figure 3 (right panel) shows the relation between the

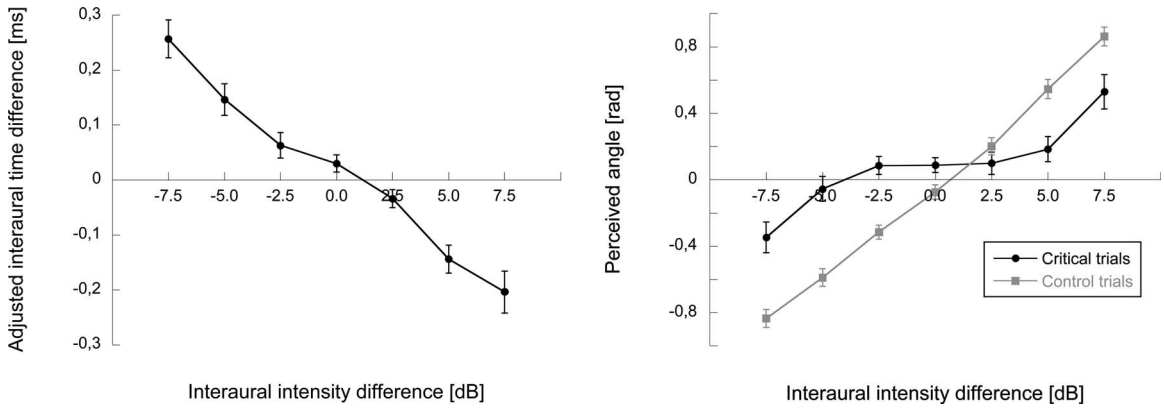


FIG. 3. Left panel: ITDs adjusted by participants in order to compensate for preset IIDs during the compensation phase of experiment 2 (error bars denote standard errors of the means). Right panel: Relation between the preset IID and perceived location during the critical trials and the control trials of the localization phase of experiment 2 (error bars denote standard errors of the means).

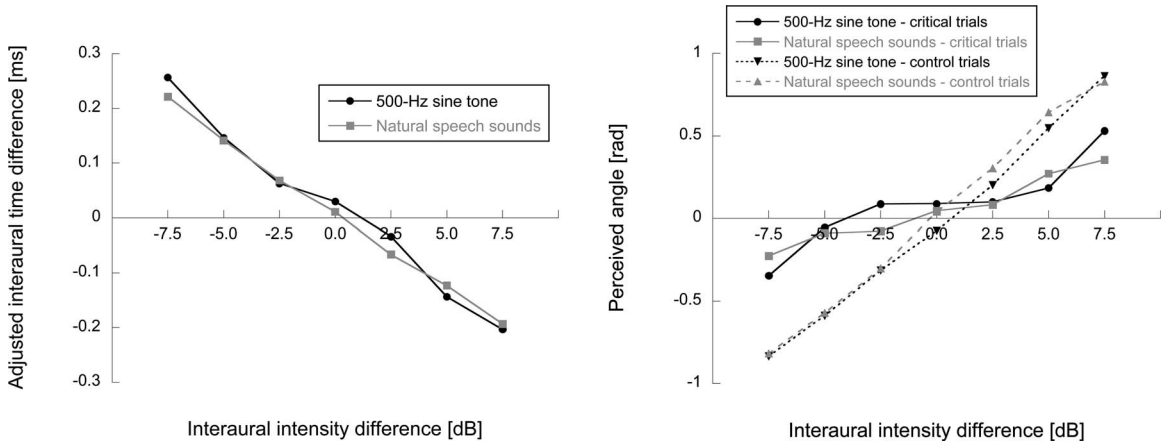


FIG. 4. Left panel: ITDs chosen to compensate for given IIDs during the compensation phase of the current experiment 2 (500-Hz sine tone) and during the compensation phase of experiment 2 of Lang and Buchner (2008) (natural speech sounds). Right panel: Relation between the preset IID and perceived location during the critical trials and the control trials of the localization phases of both experiments.

preset IID and the perceived location during the localization phase. A MANOVA for the control trials showed a significant effect of the IID on perceived sound source location [$F(6,30)=67.5$, $p<0.001$, $\eta^2=0.93$].

As in experiment 1, the most interesting analysis concerns the critical trials since an effect of the preset IID on sound source location would indicate the presence of a shift-back effect. A MANOVA for the critical trials showed that the effect of the preset IID was statistically significant [$F(6,30)=5.29$, $p=0.001$, $\eta^2=0.51$]. An analysis of the polynomial contrasts showed that the linear and the cubic components were statistically significant [$F(1,35)=20.96$, $p<0.001$, $\eta^2=0.38$ and $F(1,35)=17.12$, $p<0.001$, $\eta^2=0.33$, respectively].

C. Discussion

The central finding of experiment 2 is that, again, the shift-back effect found in experiment 2 of Lang and Buchner (2008) also emerged when 500-Hz sine tones were used instead of natural speech sounds. Even the size of the shift-back effect was almost identical in both of these experiments ($\eta^2=0.51$ and $\eta^2=0.48$, respectively).

Figure 4 shows that, in contrast to experiment 1, there was no obvious difference as to the compensation values chosen by participants between this experiment and the analogous experiment 2 of Lang and Buchner (2008) with natural speech sounds. Thus, the current experiment very nicely replicates those earlier results, showing that the shift-back effect is not tied to the use of broadband natural speech sounds.

There was, however, a more subtle difference between these experiments. In the present experiment 2, the number of trials in which the "Not enough" checkbox was chosen was clearly larger than in experiment 2 of Lang and Buchner (2008) (12.3% versus 3.1%, respectively[3]). It is not quite clear how this should be explained. Again, this may represent just random variation between experiments. However, a plausible explanation could be that the present sine tones were more difficult to localize in general than the natural speech sounds used in our previous study (e.g., see Stevens and Newman, 1936), and that the "Not enough" response category also reflects cases in which participants felt that they had not enough information about the sound source location, such that they found it impossible to adjust a "correct" value. This would also explain why the "Not Enough" checkbox was occasionally selected with small preset IIDs and even an IID of zero.

As already mentioned in Lang and Buchner, 2008, a possible problem of not presenting a reference tone that indicates the central position is that participants' internal representation of the central position might be incorrect and thus lead to a deviation from the central position during the compensation phase as compared to the localization phase where a pointing device poses a reference to the center. However, it was also noted that a systematic deviation of participants' representation of the central position seemed very implausible because preset interaural cues with an ITD of 0 μs or an IID of 0 dB had been "compensated" by values very close to zero of the other cue (see Figs. 2 and 4). Another way to test whether participants' representation of the central position is correct is a "compensation phase" in which a preset IID has to be compensated by an IID (instead of an ITD) while the ITD is 0 μs. If participants choose a mean IID of 0 dB (while the ITD is fixed at 0 μs) it is even more plausible to assume that participants' representation of the central position was correct. Experiment 3 was conducted in order to test this prediction.

IV. EXPERIMENT 3

A. Method

1. Participants

Participants were 13 female and 2 male persons, most of whom were students at the Heinrich-Heine-Universität Düsseldorf. Their age ranged from 20 to 50 years ($M=24.8$, $SD=7.3$). All participants reported normal hearing. They were paid for participating or received course credit.

2. Apparatus, stimuli, and procedure

The experiment consisted of a single phase which was similar to the compensation phases of the former experiments. A sine tone identical to that of experiments 1 and 2 was presented with one of seven preset intensity differences (−7.5, −5.0, −2.5, 0, 2.5, 5.0, or 7.5 dB). Participants were instructed to move a control element such that the tone appeared to originate from a central position. In contrast to the compensation tasks of our former experiments, the control element was not associated with the complementary interaural cue (ITD, in this case); rather, the control element regulated the preset cue (IID). ITDs were set to 0 μs during all trials.

The control element covered a range of 30 dB. In order to prevent participants from simply adjusting the control element to its middle position by visual control (i.e., choosing the middle position of the slider regardless of the position of the sound source), the range of the slider was shifted randomly on a trial-by-trial basis according to the following algorithm: The standard range was between −15 and +15 dB (identical to our previous experiments). A random number between −15 and +15 (at a resolution of 0.01) was chosen and the standard range as a whole was shifted by this value; that is, the random number was added to the low end of the range (−15 dB) and to the high end (+15 dB). Thus, the low end varied between −30 and 0 dB, and the high end varied between 0 and +30 dB while the magnitude of the range was constant at 30 dB. By applying this algorithm, it was achieved that the correct value to adjust (0 dB) could be at any position of the slider with the same probability. The mapping of the slider (low values—bottom, high values—top, or vice versa) was counterbalanced across participants. The starting position of the slider at the beginning of each trial matched the preset IID in the actual slider range. 35 trials were presented such that every preset IID appeared five times.

3. Design

The independent variable was the preset IID that was manipulated within-subject in seven steps (−7.5, −5.0, −2.5, 0, 2.5, 5.0, and 7.5 dB). Dependent variable was the IID chosen by participants to accomplish localization at the central position. A repeated-measures ANOVA was used in order to test the chosen IIDs against zero.

B. Results

Figure 5 shows the relation between the preset IIDs and the IIDs chosen by participants. A repeated-measures ANOVA revealed that the IIDs chosen by participants did not differ significantly from zero [$F(1,14)=1.75$, $p=0.207$, $\eta^2=0.11$].

C. Discussion

The IIDs chosen by participants in order to accomplish localization in the center are very close to zero. Given that the preset ITD was 0 μs in all trials, participants chose the correct value for the IIDs in order to localize the tones in the center. The control trials of experiment 1 for which the preset ITD was 0 μs show that trials with both interaural cues set to zero were located very close to the center during the localization phase (see Fig. 2—right panel). If there were a systematic deviation of participants' representation of the central position during the compensation phase as compared to the localization phase, one would expect that the IIDs chosen during the compensation phase would not lead to a localization in the center during the localization phase. This was not the case. Taken together with the finding from previous experiments that preset interaural cues with an ITD of 0 μs or an IID of 0 dB are reliably compensated by values very close to zero of the other cue (see Figs. 2 and 4), the results of experiment 3 let us confidently conclude that the shift-back effect found cannot be ascribed to an invalid internal representation of the central position during the compensation phases.

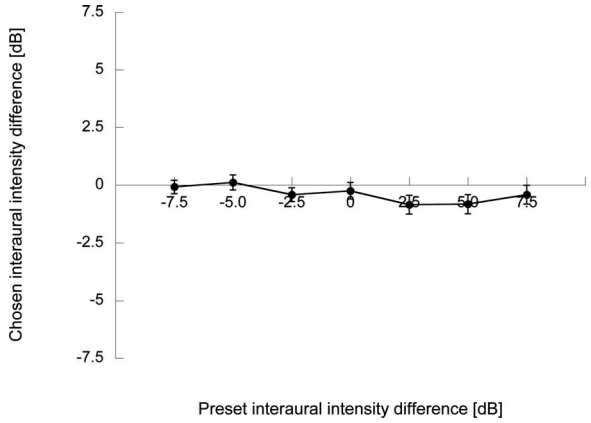


FIG. 5. Relation between preset IIDs and the IIDs chosen by participants in order to achieve a localization in the center (error bars denote standard errors of the means).

V. GENERAL DISCUSSION

The main purpose of experiments 1 and 2 was to answer the question whether the shift-back effect found with natural speech (Lang and Buchner, 2008) could be replicated with tones that had only one frequency component. The important result thus is that the shift-back effect occurs with both natural speech and 500-Hz sine tones. If there were no shift-back effect when 500-Hz sine tones were used instead of natural speech sounds, then the shift-back effect could be explained by assuming shifts of attention between different frequency ranges. However, the fact that the shift-back effect was replicated in the current experiments supports the original explanation according to which attention is shifted toward the effect of the to-be-adjusted binaural cue during the compensation phase, thereby increasing the perceptual weight of this cue above the level with which this cue affects "neutral" localization judgments. During the localization, phase attention is distributed more evenly across the binaural cues and thus, the previously adjusted cue (IIDs in experiment 1 and ITDs in experiment 2) was not large enough to compensate for the effects of the other cue on sound source localization.

Experiment 3 tested whether participants' representation of the central position underlies a systematical error if no

reference tone is present. The results suggest that this is not the case, a finding that is perfectly consistent with the fact that preset interaural cues with an ITD of 0 μs or an IID of 0 dB were reliably compensated by values very close to zero of the other cue in our previous experiments (see Figs. 2 and 4).

In more general terms, these results confirm our earlier conclusions that equivalence relations of ITDs and IIDs depend in part on states of the observer. Thus the method used to obtain equivalence relations must be taken into account when interpreting them. Specifically, relations found by setting both binaural cues into opposition must not be compared with relations found in experiments where only one cue was present at a time (such as the control trials in our experiments).

[1]The more common term "trading ratio" suggests a linear relationship between time and intensity differences. In the following, the term "trading ratio" is used either when a linear relationship is assumed or when the relation at a distinct point is reported (e.g., 80.1 μs/dB, given a time difference of 600 μs). In all other cases, the more general term "equivalence relation" is used.

[2]In the rest of the article, negative ITDs or IIDs denote that a sound was earlier or more intense, respectively, on the left channel whereas positive values indicate that it was earlier or more intense, respectively, on the right channel.

[3]Note that these values also include the trials of participants that had been excluded since they chose the checkbox for all trials of one or more preset IIDs.

Hafter, E. R., and Jeffress, L. A. (**1968**). "Two-image lateralization of tones and clicks," J. Acoust. Soc. Am. **44**, 563–569.

Lang, A.-G., and Buchner, A. (**2008**). "Relative influence of interaural time and intensity differences on lateralization is modulated by attention to one or the other cue," J. Acoust. Soc. Am. **124**, 3120–3131.

Macpherson, E. A., and Middlebrooks, J. C. (**2002**). "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," J. Acoust. Soc. Am. **111**, 2219–2236.

Stevens, S. S., and Newman, E. B. (**1936**). "The localization of actual sources of sound," Am. J. Psychol. **48**, 297–306.

Strutt, J. W. (**1907**). "On our perception of sound direction," Philos. Mag. **13**, 214–232.

Trahiotis, C., and Kappauf, W. (**1978**). "Regression interpretation of differences in time-intensity trading ratios obtained in studies of laterality using the method of adjustment," J. Acoust. Soc. Am. **64**, 1041–1047.

Whitworth, R. H., and Jeffress, L. A. (**1961**). "Time vs. intensity in the localization of tones," J. Acoust. Soc. Am. **33**, 925–929.

Young, L., and Levine, J. (**1977**). "Time-intensity trades revisited," J. Acoust. Soc. Am. **61**, 607–609.

Localization interference between components in an auditory scene

Adrian K. C. Lee
Hearing Research Center, Boston University, Boston, Massachusetts 02215 and Speech and Hearing Bioscience and Technology Program, Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts 02139

Ade Deane-Pratt
Hearing Research Center, Boston University, Boston, Massachusetts 02215

Barbara G. Shinn-Cunningham[a)]
Hearing Research Center, Boston University, Boston, Massachusetts 02215 and Speech and Hearing Bioscience and Technology Program, Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts 02139

(Received 8 May 2008; revised 24 April 2009; accepted 31 August 2009)

Some past studies suggest that when sound elements are heard as one object, the spatial cues in the component elements are integrated to determine perceived location, and that this integration is reduced when the elements are perceived in separate objects. The current study explored how object localization depends on the spatial, spectral, and temporal configurations of sound elements in an auditory scene. Localization results are interpreted in light of results from a series of previous experiments studying perceptual grouping of the same stimuli, e.g., Shinn-Cunningham *et al.* [Proc. Natl. Acad. Sci. U.S.A. **104**, 12223–12227 (2007)]. The current results suggest that the integration (pulling) of spatial information across spectrally interleaved elements is obligatory when these elements are simultaneous, even though past results show that these simultaneous sound elements are not grouped strongly into a single perceptual object. In contrast, perceptually distinct objects repel (push) each other spatially with a strength that decreases as the temporal separation between competing objects increases. These results show that the perceived location of an attended object is not easily predicted by knowledge of how sound elements contribute to the perceived spectro-temporal content of that object.
 [DOI: 10.1121/1.3238240]

PACS number(s): 43.66.Qp, 43.66.Mk, 43.66.Pn [RYL] Pages: 2543–2555

I. INTRODUCTION

In everyday life, the sound arriving at our ears is the sum of energy from multiple acoustical events in the environment, typically originating from many different sources at different locations in space. The cognitive process of interpreting the sound energy coming from different sound sources and forming the sound in the mixture into distinct perceived objects is known as auditory scene analysis or ASA (Bregman, 1990). Such objects can then be attended, allowing a listener to process an object of interest and judge its content. While it is important to be able to understand the spectro-temporal content of a signal of interest (i.e., "what" you are listening to), the spatial location of that source is also behaviorally important (i.e., "where" an auditory event comes from). For example, in a cocktail party, you not only need to able to hear your name when it is spoken (Cherry, 1953), but you also want to know the location of the person calling you.

A number of studies have investigated how ASA influences the ability to understand an attended signal (Darwin and Hukin, 1999; Freyman *et al.*, 1999; Arbogast *et al.*, 2002; Shinn-Cunningham *et al.*, 2005a). However, there are relatively few studies investigating how competing sources in a sound mixture affect localization of the perceived objects in an auditory scene. Moreover, results of these past studies show that presenting multiple sound components in a mixture can cause many different effects on sound localization.

Although some simultaneous sounds can be localized quite accurately, without strong perceptual interference between the resulting objects (Good and Gilkey, 1996; Lorenzi *et al.*, 1999; Best *et al.*, 2005), other studies suggest that simultaneous sound elements coming from different locations interfere with localization of a target element, even if the interfering elements and the target element are spectrally remote from one another (McFadden and Pasanen, 1976; Best *et al.*, 2007). Moreover, the literature on how spatial perception is affected by interactions between competing sound elements contains evidence for spatial "pulling" and "pushing" effects, as defined below (e.g., see Gardner, 1969).

Pulling (also known as "integration" or "attraction") occurs when spatial information from different sound elements is perceptually combined. Pulling causes the perceived spatial location of a target sound to be displaced toward the

a)Author to whom correspondence should be addressed. Electronic mail: shinn@cns.bu.edu

location at which the competing elements would be perceived if they were presented in isolation. Pulling has been observed, for instance, when subjects localize a source in the presence of an interfering stimulus delivered monaurally (Butler and Naunton, 1964). Another robust example of pulling is the precedence effect, in which the perceived location of a target sound closely following a preceding sound is dominated by the spatial cues in the preceding sound (see Litovsky *et al.*, 1999 for a review). A recent review of studies in which pulling occurs for sources that are spectrally remote (Best *et al.*, 2007) rekindled the idea that the degree of integration of spatial cues in different sound elements is directly affected by auditory grouping (see also Woods and Colburn, 1992).

Specifically, pulling seems to occur when sound elements are perceived as coming from the same auditory object, but this integration is reduced when grouping cues promote perceiving the spectrally remote elements in distinct auditory objects (i.e., spatial cues are perceptually integrated across only those sound elements making up a target object).

Pushing (also known as "repulsion") occurs when the perceived location of a target is displaced away from the location at which competing elements would be perceived if they were presented in isolation (Lorenzi *et al.*, 1999; Braasch and Hartung, 2002). In contrast to pulling, pushing is thought to arise when competing sounds are perceived as coming from distinct auditory objects, each of which is heard at a unique position (Best *et al.*, 2005).

To test the hypothesis that pulling occurs within objects and pushing occurs between objects, we measured the perceived laterality of auditory objects using stimuli identical to those used previously to explore the influence of spatial cues on perceived object content in a sound mixture (Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b). Briefly, when presented with a sound mixture containing a slowly repeating harmonic complex, a single target harmonic is perceived as part of the complex (Darwin and Hukin, 1997; Shinn-Cunningham *et al.*, 2007). However, if there are intervening tones that, together with the target, form an isochronous sequence of tones identical to the target, the target is typically no longer heard as part of the simultaneous harmonic complex (Darwin and Hukin, 1997; Shinn-Cunningham *et al.*, 2007). Most importantly, when the spatial cues of the target and intervening tones are manipulated, the manipulation strongly influences the perceived rhythm of the rapidly repeating tone sequence (the contribution of the target to the tone stream), but not the perceived content of the harmonic complex (not the contribution of the target to the simultaneous complex; Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b). Specifically, the tone stream is most often perceived with a galloping rhythm (the target harmonic is not part of the tone stream) when spatial cues promote (1) grouping the target with the simultaneous complex and (2) segregating the target tone and intervening harmonic tones. However, the tone stream is most often perceived with an even rhythm (the target is heard in the tone stream) when spatial cues promote (1) segregating the target tone from the complex and (2) integrating the target tone into the tone stream. Thus, spatial cues strongly affect how much the target contributes to the intervening tone stream. However, the spatial cues have only a weak effect on the perceived contribution of the target to the harmonic complex: in the presence of the tone stream, the target never strongly contributes to the complex, regardless of the spatial cues (Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b). It is worth noting that the target tone is never heard as a distinct object in these mixtures. Instead, all of these mixtures are perceived as containing only two perceptual objects: the tone stream and the harmonic complex. Manipulating the spatial cues of the sound elements simply changes the degree to which the target tone contributes to the perceived spectro-temporal content of the two objects in the scene.

The perceptual organization of mixtures of this sort (containing a rapidly repeating tone stream and a more slowly repeating harmonic complex, each of which competes for "ownership" of an ambiguous target element) is robust. If the salience of the spatial cues is reduced by adding ordinary reverberant energy, grouping results are similar, but the degree to which spatial cues modulate the perceptual contribution of the target to the tone stream is reduced (Lee and Shinn-Cunningham, 2008b). If the tone stream is changed from a simple pure tone to a complex tone containing multiple harmonics, the ambiguous target (now also a complex tone, with rich harmonic structure) contributes more to the harmonic complex, but the perceptual contribution of the target to the objects in the scene is still modulated by the spatial cues of the constituent sound elements (Lee *et al.*, 2008). If the frequency of the repeating tones vying for ownership of the target is offset from the frequency of the ambiguous target tone, the degree to which the target contributes to the tone stream decreases as the frequency disparity increases, but the same general trends are seen (i.e., spatial cues have a strong effect on the perceived content of the tone stream, but have a weaker effect on the perceived content of the harmonic complex; Lee and Shinn-Cunningham, 2008a). Thus, although how listeners group complex sound mixtures can be hard to measure precisely, multiple studies investigating mixtures like those used in the current study support the notion that there is a consistent, natural way to group these mixtures that depends on the balance of all of the various factors that affect perceptual grouping, including the spatial cues of the components in the mixture. Most importantly, for mixtures identical to those investigated here, a simple target tone is never heard strongly as part of the simultaneous harmonic complex, regardless of the spatial cues. However, the contribution of the target to the tone stream depends strongly on the spatial cues of the elements making up the mixture.

In the current study, we measured *where* subjects perceived the intervening tones and harmonic complex for stimuli identical to those used in Shinn-Cunningham *et al.* (2007). Specifically, using an interaural level difference (ILD) pointer, subjects matched the perceived laterality of either the intervening tones or the harmonic complex. Taken together with the results of our previous experiments, we find evidence for obligatory integration of spatial cues in elements presented simultaneously, even across sound elements that are not strongly perceived to be in the same au-

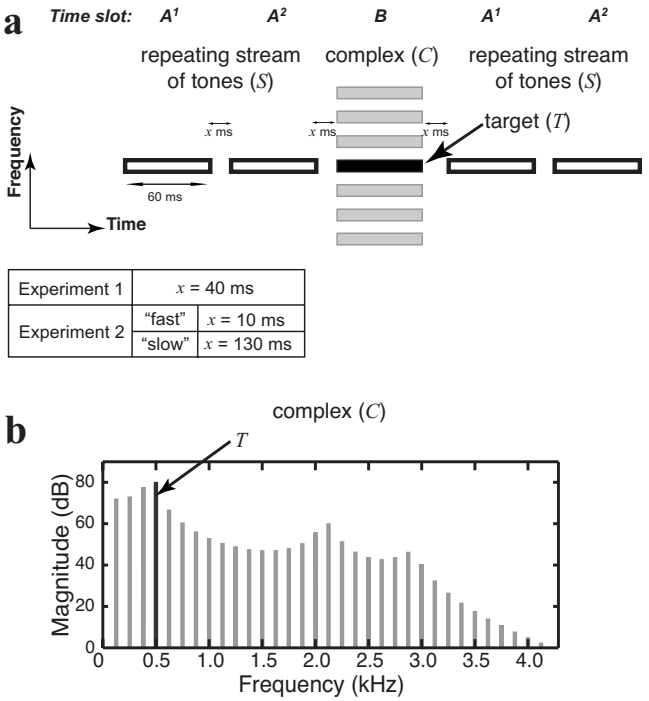


FIG. 1. (A) The two-object stimulus consists of a three-part sequence: a pair of pure tones followed by a harmonic complex (in the form of A^1A^2B). In the basic configuration, the pure tones in time slots A^1 and A^2 (S) are at 500 Hz. Time slot B is made up of two components: a target tone at 500 Hz (T) and a harmonic complex (C) with an $F0$ of 125 Hz (with the fourth harmonic 500 Hz omitted). In Experiment 1, each time slot is 100 ms in duration (60-ms-long acoustic events with 40-ms-long silent gaps between). In Experiment 2, the time slot duration is 70 ms in the "fast" block and 190 ms in the "slow" block (with silent gaps of duration 10 and 130 ms, respectively). See text for details. (B) The complex spectral envelope was shaped to sound like the vowel /ɛ/ when the target was heard as part of the complex.

ditory object. This result demonstrates a dissociation between how sound elements contribute to the perceived spectro-temporal content of objects in a scene and how the spatial cues in constituent sound elements contribute to the perceived locations of those objects. We also observe a spatial repulsion between the perceived location of competing objects. Finally, we show that the strength of the across-object repulsion decreases as the temporal separation between competing objects increases, but the across-element integration of simultaneous elements is unaffected by the temporal separation between competing objects. This final result is further evidence that how auditory elements are grouped into perceptual objects (which is strongly influenced by the temporal separation of the elements) does not always predict how spatial information is combined across elements to determine perceived object location.

II. EXPERIMENT 1: SINGLE REPETITION RATE

Stimuli consisted of a sequence of two repeating tones (S) and a harmonic complex (C) that repeated at one-third the rate of the tones. A 500-Hz tone known as the target (T) could logically belong to both the stream of tones and the harmonic complex [see Fig. 1(a)]. We manipulated the spatial content of the repeating-tone stream, the complex, and the target to explore how spatial cues influence the localization of the perceived objects. Comparison with results of grouping experiments using identical stimuli (Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b), let us explore whether there was a direct relationship between grouping and localization, as has been previously posited (Woods and Colburn, 1992; Best *et al.*, 2007). We hypothesized that if the target was strongly grouped with the simultaneous complex, the perceived location of the complex would be strongly pulled by the spatial cues of the target. However, we predicted that when the target was not heard as part of the complex, its perceived location would have little influence on the perceived location of the complex, consistent with recent results for binaural interference stimuli (Best *et al.*, 2007). We expected the repeating tone stream to be pulled by the spatial cues in the target when it was heard as part of the tone stream, but not when it was not heard as part of the tone stream. Finally, we predicted that there would be pushing between the two objects (complex and tone stream) if they were perceived in different locations (Best *et al.*, 2005).

A. Methods

1. Stimuli

The frequency of the pair of repeating tones making up the tone stream was 500 Hz [Fig. 1(a)]. The harmonic complex was filtered so that its spectral structure was vowel shaped, to enable direct comparison with our companion studies of perceptual organization using identical stimuli [Fig. 1(b); Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b]. The target was a 500-Hz tone that had the same onset/offset as the complex and that, when taken together with the repeating tones, formed an isochronous stream of identical 500-Hz tones.

The amplitudes of the target and the tones were equal, and matched the level that the fourth harmonic of the complex would have, given the spectral shaping applied to the complex. This basic pattern, a pair of repeating tones followed by the harmonic complex and target, was repeated to produce a stimulus that was perceived as two streams: an ongoing stream of tones and a repeating complex occurring at a rate one-third as rapid.

The tones, the harmonic complex, and the target were all gated with a Blackman window of 60-ms duration. There was a 40-ms-long silent gap between each tone and the simultaneous harmonic complex, creating a regular rhythmic pattern with an event occurring every 100 ms. In order to control build-up of streaming, which is known to affect perceptual grouping (Bregman, 1978; Anstis and Saida, 1985; Carlyon *et al.*, 2001; Cusack *et al.*, 2004), we kept the presentation time of these stimuli fixed at three seconds (i.e., ten repetitions of the pair-of-tones-and-complex triplet).

A number of past studies used an ILD acoustic pointer to get repeatable measures of perceived object laterality (Bernstein and Trahiotis, 1985; Trahiotis and Stern, 1989; Buell *et al.*, 1991; Heller and Trahiotis, 1996; Bernstein and Trahiotis, 2003; Best *et al.*, 2007). Therefore, we used a 200-Hz-wide band of noise, centered at 2 kHz, as an acoustic pointer, which listeners used to indicate the perceived laterality of the attended object in each trial. Subjects adjusted the

ILD of the pointer using one button to increase and another button to decrease the pointer's ILD. We used this procedure to quantify the perceived location of object whose constituent sound elements had spatial cues from pseudo-anechoic head-related transfer functions (HRTFs) measured on a KEMAR manikin at a distance of 1 m in the horizontal plane (see Shinn-Cunningham *et al.*, 2005b for details). In general, subjects did not find our HRTF-processed stimuli particularly well externalized, but they nonetheless found it intuitively easy to match the intra-cranial location of each object with the ILD pointer. Moreover, as in past studies, the ILD matches we obtained were very consistent and repeatable.

2. Task

The same physical stimuli were presented in two experimental blocks. In one block, subjects matched the perceived location of the repeated tones with the acoustic pointer. In the other block, subjects matched the perceived location of the harmonic complex. The order of stimuli was a different random sequence for each subject and each block to mitigate any learning effects.

3. Equipment

All stimuli were generated offline using MATLAB software (Mathworks Inc.). Sources were processed to have spatial cues consistent with a source from a position straight ahead (0° azimuth), 45° to the left, or 45° to the right of the listener.

Digital stimuli were generated at a sampling rate of 25 kHz and sent to Tucker-Davis Technologies (TDT) hardware for D/A conversion and attenuation before presentation over headphones (Etymotic ER-1 insert earphones). Presentation of the stimuli was controlled by a personal computer, which selected the stimulus to play on a given trial. A different random attenuation level (0–14 dB) was applied to both the stimulus and the acoustic pointer in each trial in order to minimize any influence of presentation level on localization. Subjects were seated in a sound-treated booth and responded via a button-box (TDT Bbox), which was directly connected to the hardware. All signals were presented at a listener controlled, comfortable level (maximum value 80 dB sound pressure level).

4. Subjects

Nine subjects (four male, five female, aged 18–31) took part in the experiment. All participants had pure-tone thresholds in both ears within 20 dB of normal-hearing thresholds at octave frequencies between 250 and 8000 Hz, and within 15 dB of normal-hearing thresholds at 500 Hz. All subjects gave written informed consent to participate in the study, as overseen by the Boston University Charles River Campus Institutional Review Board and the Committee On the Use of Humans as Experimental Subjects at the Massachusetts Institute of Technology.

B. Procedures

1. Training

At the beginning of each experimental block, all listeners received 15 min of practice to familiarize themselves with the experimental procedures and task, which were identical to those of the main experiment (described below). During these practice sessions, subjects were encouraged to explore the full range of acoustic pointer positions they could achieve, and diagrams were presented on screen to help emphasize the difference between the repeating tones and the harmonic complex. No feedback was provided either during training or during the main experiment.

2. Matching procedure

Each trial began with a presentation of the 3-s-long stimulus. This was followed by a 3-s-long presentation of the acoustic pointer, during which subjects could adjust its ILD. A right button press caused the ILD to increase by one step while a left button press caused it to decrease (achieved by symmetrically adjusting the level to the right ear upward and the level to the left ear downward by the same amount). Updates occurred at a rate of 25 kHz. The constant step-size was set to be very small ($|\Delta| = 1.5 \times 10^{-3}$ dB), so that listeners perceived, in real time, an essentially continuous sound image moving along the intracranial axis as they adjusted the pointer ILD.

Presentations of the stimulus and pointer alternated every 3 s until the subject was satisfied that the pointer laterality matched the perceived laterality of the attended object. To indicate their satisfaction, they pressed a third button, which caused the current pointer ILD to be stored and the next trial to be initiated. The initial pointer ILD was set to a random value (between −20 and +20 dB) at the start of each trial. Typically, subjects cycled through three to four iterations of the listening-matching sequence for each trial before signaling satisfaction with their response.

3. Blocking

Each block of the experiment included seven single-object conditions that served as controls (see left panels of Fig. 2). In three of these conditions, the target and the object to be localized (either the complex or the repeating tones) were both from the same location, either at 0°, 45° (to the right), or −45° (to the left). In the four other single-object conditions, the target and the object to be localized were from different locations (but there were no other competing objects). Seven two-object conditions (which were identical in the "match-tones" and "match-complex" blocks) were intermingled with the appropriate single-object conditions in each block (right panel in Fig. 2). In these conditions, the complex always originated from 0° azimuth. In one two-object control condition, the target and the repeating tones were co-located with the complex. The other six conditions consisted of two conditions in which only the target came from the side (either left or right) and the complex and repeating tones were straight ahead; two with only the repeating tones coming from the side (either left or right) and the complex and target from straight ahead; and two with the

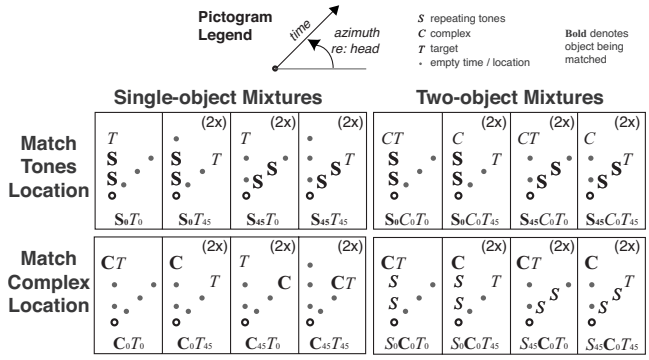


FIG. 2. Summary of the spatial configurations tested (mirror-symmetric versions of the three right-most conditions of each group were presented, for a total of seven stimuli of each type). Single-object conditions are shown on the left for the match-tones (top) and match-complex conditions (bottom). Two-object conditions, presented in both the match-tones and match-complex blocks, are shown on the right. The radial dimension in each diagram denotes time, while the azimuthal angle of each component relative to the listener is denoted by the angle relative to the top-down view of the head.

repeating tones and the target both coming from the same side (either left or right) and the complex from ahead.

Because conditions were mirror symmetric and there were no significant differences in either the main effect of side of presentation (tones: $F_{1,7}=0.449$; complex: $F_{1,7}=0.757$) or the interaction between side of presentation and condition (tones: $F_{7,52}=0.494$; complex: $F_{3,20}=0.534$), results from left/right symmetric conditions were combined in all subsequent analysis. The resulting configurations are denoted by the shorthand $S_xC_yT_z$, where x, y, and z denote the locations of the stream of tones (S), complex (C), and target (T) and can either be 0 for center (components from 0° azimuth) or 45 for side (components from either ±45° azimuth). Bold font highlights the component of the mixture that listeners were asked to match in a given condition. This leads to four unique configurations for the single-object match-tones conditions ($\mathbf{S}_0T_0$, $\mathbf{S}_0T_{45}$, $\mathbf{S}_{45}T_0$, $\mathbf{S}_{45}T_{45}$; see Fig. 2, top left panel), four configurations for the single-object, match-complex conditions ($\mathbf{C}_0T_0$, $\mathbf{C}_0T_{45}$, $\mathbf{C}_{45}T_0$, $\mathbf{C}_{45}T_{45}$; see Fig. 2, bottom left panel), and four two-object configurations that were presented in both match complex ($S_0\mathbf{C}_0T_0$, $S_0\mathbf{C}_0T_{45}$, $S_{45}\mathbf{C}_0T_0$, and $S_{45}\mathbf{C}_0T_{45}$) and match-tone conditions ($\mathbf{S}_0C_0T_0$, $\mathbf{S}_0C_0T_{45}$, $\mathbf{S}_{45}C_0T_0$, and $\mathbf{S}_{45}C_0T_{45}$; see Fig. 2, right panel).

Each subject completed two experimental blocks, each consisting of 8 repetitions of each of the 14 stimuli in random order, for a total of 224 trials per block. The order of the blocks was counter-balanced across sessions and listeners. Each session lasted no longer than 1.5 h.

C. Results

In all data analysis, results from mirror-symmetric configurations were combined by reversing the sign of the ILD match for configurations with elements to the left and then averaging these values with the corresponding results for configurations with elements to the right. Table I summarizes results of all the statistical tests performed on results from Experiment 1, discussed in detail below.

1. Single-object mixtures

a. Tones. When tone stream and target were both from 0° [$\mathbf{S}_0T_0$: left-most data point in Fig. 3(a)], the mean ILD was near zero (i.e., when both target and tones were from straight ahead, the tones were perceived at midline). Results were similar when the target was shifted to the side and the tones remained in the center [$\mathbf{S}_0T_{45}$; compare the first and second data points from the left in Fig. 3(a)]. When tones and target were both to the side [$\mathbf{S}_{45}T_{45}$; right-most data point in Fig. 3(a)], the mean ILD was large and pointing to the expected side. When the tones were from the side and the target was from the center [$\mathbf{S}_{45}T_0$: third data point from left in Fig. 3(a)], the perceived location of the tones was shifted toward midline compared to when both tones and target were to the side [compare the two right-most data points in Fig.

TABLE I. Summary of paired-sample *t*-test and Wilcoxon signed-rank tests (comparisons shown in italics) performed on the group mean of Experiment 1 after collapsing across left-right symmetric configurations (see also Fig. 3). *Post-hoc* adjusted significance levels, using the Dunn–Sidak factors, are reported here and are denoted by the subscript DS. Ellipses denote comparisons for which both conditions may have been affected by the limited response range and whose significance was therefore not tested.[1]

		Effect of target location		Effect of competing object		Effect of tones location	
		Conditions	Significance	Conditions	Significance	Conditions	Significance
Single-object	Tones	$\mathbf{S}_0T_{45}-\mathbf{S}_0T_0$	$t_8=-1.327$ $p_{DS,2}=0.113$				
		$\mathbf{S}_{45}T_{45}-\mathbf{S}_{45}T_0$	***Z=−3.4187*** ***$p_{DS,2}=0.001$***				
	Complex	$\mathbf{C}_0T_{45}-\mathbf{C}_0T_0$	$\mathbf{t_8=-3.267}$ $\mathbf{p_{DS,2}=0.023}$				
		$\mathbf{C}_{45}T_{45}-\mathbf{C}_{45}T_0$	⋯ ⋯				
Two-object	Tones			$\mathbf{S}_0C_0T_{45}-\mathbf{S}_0T_{45}$	$t_{17}=-2.211$ $p_{DS,3}=0.118$		
		$\mathbf{S}_0C_0T_{45}-\mathbf{S}_0C_0T_0$	$t_8=2.689$ $p_{DS,2}=0.054$	$\mathbf{S}_{45}C_0T_0-\mathbf{S}_{45}T_0$	$Z=2.025$ $p_{DS,3}=0.123$		
		$\mathbf{S}_{45}C_0T_{45}-\mathbf{S}_{45}C_0T_0$	⋯ ⋯	$\mathbf{S}_{45}C_0T_{45}-\mathbf{S}_{45}T_{45}$	⋯ ⋯		
	Complex			$S_0\mathbf{C}_0T_{45}-\mathbf{C}_0T_{45}$	$t_{17}=2.144$ $p_{DS,3}=0.134$		
		$S_0\mathbf{C}_0T_{45}-S_0\mathbf{C}_0T_0$	$\mathbf{t_8=-3.638}$ $\mathbf{p_{DS,2}=0.013}$	$S_{45}\mathbf{C}_0T_0-\mathbf{C}_0T_0$	$\mathbf{t_{17}=-5.347}$ $\mathbf{p_{DS,3}=0.002}$	$S_{45}\mathbf{C}_0T_0-S_0\mathbf{C}_0T_0$	$\mathbf{t_8=4.321}$ $\mathbf{p_{DS,2}=0.005}$
		$S_{45}\mathbf{C}_0T_{45}-S_{45}\mathbf{C}_0T_0$	$\mathbf{t_{17}=-3.358}$ $\mathbf{p_{DS,2}=0.007}$	$S_{45}\mathbf{C}_0T_{45}-\mathbf{C}_0T_{45}$	$\mathbf{t_{17}=-5.184}$ $\mathbf{p_{DS,3}=0.001}$	$S_{45}\mathbf{C}_0T_{45}-S_0\mathbf{C}_0T_{45}$	$\mathbf{t_{17}=5.349}$ $\mathbf{p_{DS,2}<0.001}$

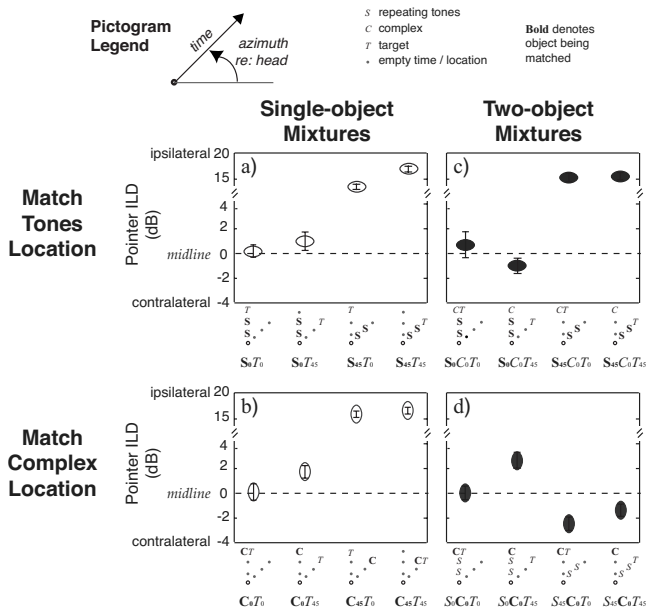


FIG. 3. Across-subject average of the matched ILD for all conditions in Experiment 1, collapsed across mirror-symmetric conditions (all but the left-most condition). Matches to the tones are denoted by horizontal ellipses (top panels) and matches to the complexes are denoted by vertical ellipses (bottom panels). Open symbols represent single-object conditions [(A) and (B)] and filled symbols represent two-object conditions [(C) and (D)]. Error bars show the standard error of the mean across subjects.

3(a); this effect was significant according to a Wilcoxon signed-rank test, $p<0.05$[1]; see comparison of $\mathbf{S}_{45}T_0$ and $\mathbf{S}_{45}T_{45}$ in Table I].

b. Complex. When the complex and target were in the same location, results are as expected: near zero for $\mathbf{C}_0T_0$ [left-most data point in Fig. 3(b)] and large and toward the expected side for $\mathbf{C}_{45}T_{45}$ [right-most data point in Fig. 3(b)]. When the target was from the side and the complex was from the center ($\mathbf{C}_0T_{45}$), the perceived location of the complex was pulled toward the side of the target [compare the two left-most data points in Fig. 3(b), $p<0.05$ for the comparison of $\mathbf{C}_0T_0$ and $\mathbf{C}_0T_{45}$ in Table I]. When the complex was simulated from the side, its perceived location was far off to the expected side, both when the target was in front and when the target was to the side ($\mathbf{C}_{45}T_0$ and $\mathbf{C}_{45}T_{45}$).[1]

2. Two-object mixtures

a. Tones. When the tone stream, the target, and the complex were all from the center [$\mathbf{S}_0C_0T_0$: leftmost data point Fig. 3(c)], the judged tone location was close to zero, as expected. There was a trend for the target location to affect the perceived location of the tones, but this trend did not reach statistical significance. Specifically, when the target was to the side and all of the other components were straight ahead ($\mathbf{S}_0C_0T_{45}$), there was a trend for the perceived tone stream location to be displaced slightly away from midline, away from the side of the target, compared to when all components were from in front [compare two left-most data points in Fig. 3(c); $p=0.054$, as shown in Table I]. When the tones and target were to the side and the complex was straight ahead ($\mathbf{S}_{45}C_0T_{45}$: right-most data point), the tones were heard far to expected side. Similarly, when the tones were to the side but the target and complex were from straight ahead, the perceived tone stream location was far to the side in the expected direction [$\mathbf{S}_{45}C_0T_0$: third data point from the left in Fig. 3(c)].[1]

The effect of adding the complex to the sound mixture can be discerned by comparing single-object and two-object judgments [corresponding open and filled horizontal ellipses in Figs. 3(a) and 3(c), respectively]. These comparisons give no strong evidence for an effect of the complex on the perceived location of the tones. When all components (target, tones, and complex) were from the front, there was no effect of adding the complex: the tones continue to be heard from midline [$\mathbf{S}_0T_0$ versus $\mathbf{S}_0C_0T_0$; compare left-most data points in Figs. 3(a) and 3(c)]. Adding the complex from in front had no statistically significant effect on the perceived location of the tones either when the tones were in front and the target was to the side [$\mathbf{S}_0T_{45}$ versus $\mathbf{S}_0C_0T_{45}$; the second data points from the left in Figs. 3(a) and 3(c), respectively] or when the tones were from the side and the target was from in front [$\mathbf{S}_{45}T_0$ versus $\mathbf{S}_{45}C_0T_0$; the third data points from the left in Figs. 3(a) and 3(c), respectively]. (When the tone stream and the target were to the side, responses may have been affected by the response range, so no statistical tests were performed.[1])

b. Complex. As expected, when the tones, the target, and the complex were all from the center, the perceived location of complex was near zero [$S_0\mathbf{C}_0T_0$: see left-most data point in Fig. 3(d)]. The perceived location of the complex was influenced by the location of the tones in the two-object mixtures. When the complex and tones were in front and the target was to the side, the perceived location of the complex was displaced toward the side of the target compared to when all three sound elements were from in front [compare the two left-most data points in Fig. 3(d); $S_0\mathbf{C}_0T_0$ and $S_0\mathbf{C}_0T_{45}$ differ significantly, with $p<0.05$, in Table I]. When the complex was in front and the tones were to the side, the perceived location of the complex also was displaced towards the side of the target when the target was moved from midline [compare the two right-most data points in Fig. 3(d); $S_{45}\mathbf{C}_0T_0$ and $S_{45}\mathbf{C}_0T_{45}$ differ significantly, with $p<0.05$, in Table I]. Thus, the effect of moving the target location was to shift the perceived location of the complex in the direction of the target in the two-mixture conditions ($S_0\mathbf{C}_0T_{45}$ versus $S_0\mathbf{C}_0T_0$ and $S_{45}\mathbf{C}_0T_{45}$ versus $S_{45}\mathbf{C}_0T_0$).

In addition to being affected by the location of the target, the perceived location of the complex could be influenced by the simple presence of the tones. When the complex and target were from the center and the tones were from the side, the perceived location of the complex was displaced from midline, away from the tones [compare the third data point from the left in Fig. 3(d) with the left-most data point in Fig. 3(b); $S_{45}\mathbf{C}_0T_0$ and $\mathbf{C}_0T_0$ differ significantly, with $p<0.05$, in Table I]. Similarly, when the complex was from the center and the target and tones were from the side, the perceived location of the complex was displaced away from the tones compared to the perceived location of the complex without the tones present [compare the right-most data point in Fig.

3(d) with the second data point from the left in Fig. 3(b); $S_{45}\mathbf{C}_0T_{45}$ and $\mathbf{C}_0T_{45}$ differ significantly, with $p<0.05$, in Table I].

Finally, the location of the competing tones also influenced the perceived location of the complex when comparing two-object conditions. Changing the location of the tones from the center to the side caused the perceived location of the complex to be displaced away from midline into the hemifield opposite the location of the tones, both when the complex and target were in the center [$S_0\mathbf{C}_0T_0$ versus $S_{45}\mathbf{C}_0T_0$; compare first and third data points in Fig. 3(d)] and when the complex was in the center and the target was to the side [$S_0\mathbf{C}_0T_{45}$ versus $S_{45}\mathbf{C}_0T_{45}$; compare the second and fourth data points from the left in Fig. 3(d)]. In both of these cases, these results, which are consistent with the competing tones repelling the perceived location of the complex, were statistically significant ($p<0.05$ for both comparisons in Table I).

D. Discussion

1. Single-object mixtures

a. Tones. In a single-object mixture, the location of a target at midline pulled the perceived location of a tone stream ($\mathbf{S}_{45}T_0$ was closer to midline than $\mathbf{S}_{45}T_{45}$). Thus, in the absence of any competing object, there can be across-time integration of spatial cues that affects the perceived location of the repeated tones. This effect is consistent with the well-known phenomenon of "binaural sluggishness" (Grantham and Wightman, 1978; Culling and Summerfield, 1998; Culling and Colburn, 2000), which is thought of as an obligatory across-time integration of spatial cues. Such sluggishness should depend strongly on the repetition rate of the stimuli, with more integration at faster rates (larger pulling from the target) and less at slower rates. In Experiment 2, we directly tested this hypothesis by comparing localization of the tones with the target at two different repetition rates.

b. Complex. The perceived location of the complex presented without the tone stream tended to be pulled toward the location of the simultaneous target ($\mathbf{C}_0T_{45}$ was pulled from midline toward the side of the target compared to $\mathbf{C}_0T_0$). Because the pulling of the complex by the target depends only on integration of simultaneously presented elements, this pulling should not be affected by repetition rate. We tested this hypothesis in Experiment 2.

2. Two-object mixtures

Just as the perceived location of the complex was pulled by the target in single-object conditions, the complex was pulled toward the location of the target when the tone stream was present in the mixture. In contrast, the target sometimes pulled the perceived location of the tones in single-object mixtures, but there was no evidence for across-time integration of the target and the tones when the complex was present. Indeed, in the two-object mixtures with the complex and tones both from in front, the perceived location of the tones had a tendency to be displaced away from the side of the target, rather than pulled toward the target. Thus, results suggest the target always pulls the perceived location of the simultaneous complex, but that the target spatial information is not integrated with the perceived location of the tones when there is a simultaneous complex present in the mixture. Given these results, we expected the manipulations of the repetition rate undertaken in Experiment 2 to have little effect on how strongly the target pulled either the perceived location of the complex (which presented simultaneously with the target) or the perceived location of the tones (which was never pulled by the target in the two-object mixtures).

When there were two competing objects in the mixture, there was a tendency for the competing objects to repel each other. Specifically, adding tones from the side caused a significant displacement of the complex away from the side of the tones ($S_{45}\mathbf{C}_0T_0$ versus $\mathbf{C}_0T_0$ and $S_{45}\mathbf{C}_0T_{45}$ versus $\mathbf{C}_0T_{45}$). Similarly, moving the tones to the side caused the perceived location of the complex to be displaced away from side of the tones, as if the tones repelled the complex ($S_0\mathbf{C}_0T_0$ versus $S_{45}\mathbf{C}_0T_0$ and $S_0\mathbf{C}_0T_{45}$ versus $S_{45}\mathbf{C}_0T_{45}$). Across-object repulsion could also explain the trend for the perceived location of the tones to be displaced away from the side of the target in condition $\mathbf{S}_0C_0T_{45}$. In the limit, if the tones and complex are separated by a large inter-stimulus interval, any across-object repulsion must disappear. Thus, we hypothesized that repulsion would be stronger when the repetition rate was faster and weaker when the rate was slower, an idea tested in Experiment 2.

III. EXPERIMENT 2: VARYING REPETITION RATES

In this experiment, three hypotheses were tested:

(1) In the tones-only conditions, the strength of the pulling of the tone stream by the target will increase with increasing repetition rate.
(2) In the complex-only conditions, the pulling of the target on the complex will be independent of the repetition rate.
(3) In the two-object conditions, the strength of across-object repulsion will increase with increasing repetition rate.

A. Methods and procedures

Stimuli were identical to those in Experiment 1 except that the length of the silent gap between each tone and complex varied [Fig. 1(a)]. In the "fast" block of the experiment, the silent gap was 10 ms and an acoustic event occurred every 70 ms. In the "slow" block of the experiment, the silent gap was set at 130 ms, with events every 190 ms. In both experimental blocks, presentation time was fixed at 3 s. As a result, stimuli in the fast block consisted of 14 repetitions of the repeating-tone-complex triplet, while in the slow block it consisted of five repetitions.

Nine subjects (four male, five female, aged 18–31) took part in this experiment. Eight out of these nine subjects also participated in Experiment 1. The training and matching procedures were identical to those used in Experiment 1. Each subject completed four experimental blocks (localization of the repeating tones and the complex at two different rates) on two separate days. On any given day, each subject completed two blocks of tone stream localization and two blocks of

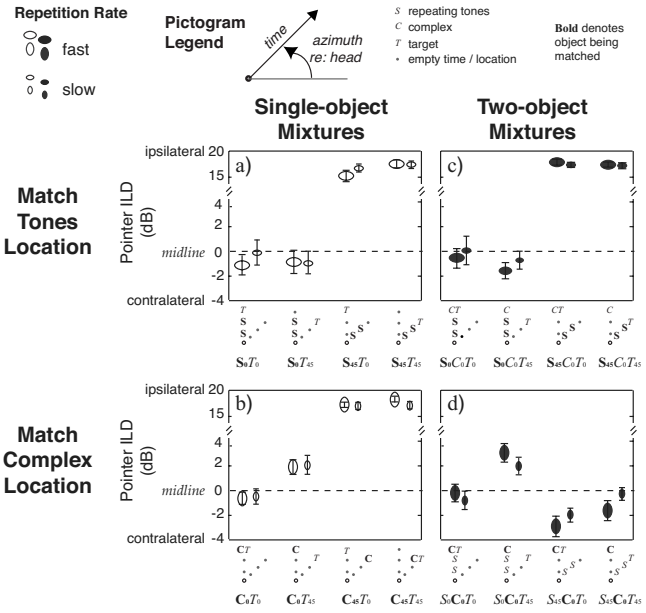


FIG. 4. Across-subject average of the matched ILD for all conditions in Experiment 2 collapsed across mirror-symmetric conditions. Matches to the tones are denoted by horizontal ellipses (top panels) and matches to the complexes are denoted by vertical ellipses (bottom panels). Open symbols represent single-object conditions [(A) and (B)] and filled symbols represent two-object conditions [(C) and (D)]. Big symbols represent the "fast" repetition rate and small symbols represent the "slow" repetition rate. Error bars show the standard error of the mean across subjects.

complex localization, one each for slow and fast repetition rates. The order of the stimulus rate and the order of the task were counter-balanced across subjects.

B. Results

1. Single-object mixtures

a. Tones. In general, repetition rate had little effect on the localization of the tones [compared the bigger and the smaller markers of each condition in Fig. 4(a)]. As expected, when tones and target were both from 0° [$\mathbf{S}_0T_0$; left-most pair of data points in Fig. 4(a)], the mean ILD was near zero, independent of the repetition rate. When the target was from the side and the repeated tones from center, the tones were still perceived near center [$\mathbf{S}_0T_{45}$; second pair of data points from left in Fig. 4(b)]. A two-way, repeated-measures analysis of variance (ANOVA) was conducted on the mean data exploring the effect of repetition rate fast versus slow and target location ($\mathbf{S}_0T_0$ versus $\mathbf{S}_0T_{45}$) on localization judgments when the repeating tones were from center. Neither of the main effects nor their interaction was statistically significant [compare the two left-most pairs of data points in Fig. 4(a); see first row of Table II].

When tones and target were both to the side [$\mathbf{S}_{45}T_{45}$: right-most pair of data points in Fig. 4(a)], the mean ILD was large and to the expected side, independent of repetition rate. When the repeating tones were to the side and the target was in front, the perceived laterality of the tones was also far to the expected side [$\mathbf{S}_{45}T_0$: third pair of data points from left in Fig. 4(a); because the response range may have affected these results, no statistical tests were performed to compare single-object tone conditions[1]].

b. Complex. The perceived location of the complex was as expected in the control conditions. When the complex and target were both from midline [$\mathbf{C}_0T_0$; left-most pair of data points in Fig. 4(b)], the ILD was near zero. When the complex and target were both to the side [$\mathbf{C}_{45}T_{45}$: right-most pair of data points in Fig. 4(b)], the ILD was large, and in the expected direction.

Consistent with results of Experiment 1, the target tended to pull the perceived location of the midline complex. When the complex was from in front and the target was to the side, judgments were displaced from midline toward the target side [$\mathbf{C}_0T_{45}$; pair of data points second from the left in Fig. 4(b)]. When the complex was from the side, judgments were far to the side for all matches [see the two right-most pairs of data points in Fig. 4(b)].

Consistent with our hypothesis, repetition rate had little effect on the localization of the complex, independent of the exact spatial configuration of the target and complex. This observation was supported by a two-way, repeated-measures ANOVAs with factors of target location and repetition rate (fast and slow) for the complex coming from the center: the main effect of target location was significant, but neither the effect of repetition rate nor the two-way interaction was statistically significant [the left-most pair of data points is lower than the pair of data points second from the left in Fig. 4(b), but within each pair, the points are similarly valued; see third line of Table II]. (When the complex was from the side, results may have been affected by the response range, so statistical tests were not performed.[1])

TABLE II. Summary of the two-way repeated ANOVA tests performed on the group mean of the single-object conditions in Experiment 2 after collapsing across left-right symmetric configurations (see also Fig. 4). Ellipses denote comparisons that were excluded, as results in both conditions may have been artificially limited by the response range.[1]

		Single-object (two-way, repeated ANOVA) Main Effect					
	Conditions	Target location		Rate		Interaction	
Tones	$\mathbf{S}_0T_{45}$–$\mathbf{S}_0T_0$	$F_{1,8}=0.092$	$p=0.348$	$F_{1,8}=3.487$	$p=0.099$	$F_{1,8}=0.929$	$p=0.363$
	$\mathbf{S}_{45}T_{45}$–$\mathbf{S}_{45}T_0$	…	…	…	…	…	…
Complex	$\mathbf{C}_0T_{45}$–$\mathbf{C}_0T_0$	**$\mathbf{F_{1,8}=16.987}$**	**$\mathbf{p=0.003}$**	$F_{1,8}=0.16$	$p=0.902$	$F_{1,8}=0.027$	$p=0.874$
	$\mathbf{C}_{45}T_{45}$–$\mathbf{C}_{45}T_0$	…	…	…	…	…	…

TABLE III. Summary of the one-tailed Wilcoxon signed-rank tests of the effect of repetition rate on the group mean of the two-object conditions in Experiment 2 (see also Fig. 4). Ellipses denote comparisons that were excluded, as results in both conditions may have been artificially limited by the response range.[1]

Two-object (Wilcoxon signed-rank tests on rate of repetition)				
Conditions	Tones	$\mathbf{S_0C_0T_{45}}$	**Z=1.677**	**p=0.047**
		$\mathbf{S_{45}}C_0T_0$	⋯	⋯
		$\mathbf{S_{45}}C_0\mathbf{T_{45}}$	⋯	⋯
	Complex	$\mathbf{S_0C_0T_{45}}$	**Z=2.286**	**p=0.011**
		$S_{45}\mathbf{C_0}T_0$	$Z=1.633$	$p=0.051$
		$\mathbf{S_{45}C_0T_{45}}$	**Z=2.112**	**p=0.035**

2. Two-object mixtures

a. Tones. When the tone stream, target, and complex were all from the center ($\mathbf{S}_0C_0T_0$), the mean ILD for the tone stream match was near zero, independent of repetition rate, as expected [see the left-most pair of data points in Fig. 4(c)]. From the results in Experiment 1, we expected any across-object repulsion to be evident only when the competing tones and complex were perceived in different locations and the responses were not far to the side (and thus not affected by a response ceiling effect[1]). Therefore, we only expected to see an effect of rate on the repulsion of the tones in configurations $\mathbf{S}_0C_0T_{45}$. A Wilcoxon signed-rank test was performed to test for the effect of rate in condition $\mathbf{S}_0C_0T_{45}$.[2] There was a significant effect of the repetition rate on the perceived location of the tones, consistent with there being repulsion of the tones that was significantly smaller for the slower repetition rate in condition $\mathbf{S}_0C_0T_{45}$ [see the pair of data points second from the left in Fig. 4(c) and the first line of Table III].

b. Complex. As expected, when the tones, the target, and the complex were all from the center ($S_0\mathbf{C}_0T_0$), the mean ILD for complex localization was near zero, independent of repetition rate. However, rate affected localization of the complex. In particular, spatial repulsion was consistently stronger for the faster repetition rate than for the slower repetition rate: within each of the three right-most pairs of data points in Fig. 4(d), the right data point in each pair is closer to the perceived location of the complex without the tones present [shown by the two left-most pairs of matches in Fig. 4(b)] than the left data point in the pair. Moreover, all of these judgments of the perceived location of the complex are displaced away from the perceived location of the tones in that mixture [shown by the corresponding results in Fig. 4(c)]. Using three separate Wilcoxon signed-rank tests, the repulsion of the harmonic complex was found to be significantly smaller for the slower repetition rate in conditions $S_0\mathbf{C}_0T_{45}$ [the pair of data points second from the left in Fig. 4(d); see fourth line of Table III] and $S_{45}\mathbf{C}_0T_{45}$ [the pair of data points third from the left in Fig. 4(d); see final line of Table III]. Although the effect of repetition rate on the localization of the complex failed to reach statistical significance in condition $S_{45}\mathbf{C}_0T_0$, there was a trend for repulsion by the tones to be weaker at the slower repetition rate even in this condition [right-most pair of data points in Fig. 4(d); in fifth line of Table III, $p=0.051$].

C. Discussion

1. Single-object mixtures

a. Tones. In the single-object conditions, there was little evidence for an effect of the target location on the perceived location of the tones. These results suggest that there is relatively little integration of the target spatial cues when judging the location of the repeated tones, even at the fastest repetition rate. This result is interesting, especially in light of the fact that listeners strongly perceive the target as part of the repeated tone object in related experiments investigating what objects listeners perceived in these kind of sound mixtures (Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b).

b. Complex. When listeners judged the laterality of the complex in single-object configurations, the target significantly pulled the perceived location of the complex when the complex originated from the center. This pulling was not significantly influenced by the rate of repetition, consistent with our hypothesis. These spatial judgments suggest that there is an obligatory integration of the spatial information in the target with the spatial cues in the simultaneously present complex, and are consistent with the fact that the spatially displaced target is heard as part of the complex when there is no other object competing for ownership of the target (Darwin and Hukin, 1997; Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b).

2. Two-object mixtures

Across-object spatial repulsion was generally stronger at the faster repetition rate and weaker when the competing objects were more separated in time, consistent with our hypothesis. This effect of repetition rate was found to be statistically significant both when localizing the tones (in condition $\mathbf{S}_0C_0T_{45}$) and when localizing the harmonic complex ($S_0\mathbf{C}_0T_{45}$ and $S_{45}\mathbf{C}_0T_{45}$; there was a trend for repulsion to be weaker at the slower repetition rate in condition $S_{45}\mathbf{C}_0T_0$).

In two-object mixtures, the perceived location of the complex depends both upon pulling by the simultaneous target and repulsion by the competing tone stream. For instance, in condition $\mathbf{C}_0T_{45}$, the side target pulls the perceived location of the complex away from midline. Adding tones from midline to the mixture causes the complex and tones to repel one another, so that the complex is heard even farther to the side of the target than when the tones are not present ($S_0\mathbf{C}_0T_{45}$ versus $\mathbf{C}_0T_{45}$). As noted above, the repulsion between the complex and the tones decreases as the temporal separation between complex and tones increases, as expected. These observations highlight the fact that both pulling and pushing can occur in the same conditions.

3. Build-up of streaming

Many studies have shown that when listeners hear a repeating sequence of elements, the way in which the listeners perceptually organize the sound mixture changes, or "builds up," over time (Bregman, 1978; Anstis and Saida, 1985; Carlyon *et al.*, 2001; Cusack *et al.*, 2004). To the extent that any build-up of streaming depends on the number of presentations, build-up should be greater in the fast block than in the

slow block. Conversely, if build-up of streaming depends on absolute time rather than the number of stimulus presentation, build up should be similar in the two blocks. Further work is necessary to investigate how streaming build-up may influence the perceived location of objects in an auditory scene. However, the current results show that across-object interactions influence where listeners perceive objects in a complex scene.

IV. GENERAL DISCUSSION

A. Pulling (integration)

Spatial information in the target influenced localization of both across-time objects (the tone stream) and objects grouped across frequency (the harmonic complex) when there were no competing objects in the mixture. The target weakly pulled the tone stream location, having an observable influence only when the tones were from the side and the target was from the center. Moreover, this pulling did not depend significantly on repetition rate. This suggests that binaural sluggishness does not cause a strong, obligatory temporal integration of spatial cues across time for the tones and target used in these experiments. However, such effects might arise in similar experiments if the temporal gap between events was smaller (or, equivalently, if the repetition rate was greater).

In both experiments, in single-object conditions containing the target and complex, the target significantly pulled the complex when the complex came from the center. This result shows that the target spatial information was integrated with the spatial information in the simultaneously presented complex in single-object conditions.

When there are two objects present in the scene, the target spatial cues did not ever pull the perceived location of the tone stream significantly, but always pulled the perceived location of the complex. There was no strong effect of repetition rate on the pulling of the complex by the target. As noted above, this makes intuitive sense, given that the complex and the target are simultaneous, suggesting that the integration of their spatial information should be independent of any temporal parameters.

The observed obligatory integration of the target spatial cues with the complex in the presence of the tones is surprising in light of past studies of perceptual organization of auditory mixtures like those used here. Specifically, when listeners are asked whether the target contributes to what elements the harmonic complex contains, the target usually is not heard as part of the complex for these mixtures (Darwin and Hukin, 1997; Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b). Conversely, the target contributes to the tone stream in some cases, but not others, depending on the spatial cues in these stimuli (Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b). In contrast, we find that the target, which is *never* heard strongly as part of the complex, *always* contributes to the perceived location of the complex. Moreover, the target never contributes to the perceived location of the tones in the two-object mixtures, regardless of the spatial configuration of the elements; however, the spatial configuration has a dramatic effect on whether or not the target is heard as part of the tone stream.

These results are interesting when taken in conjunction with results of past studies that suggest that the perceived location of an object depends on integrating spatial cues contained only in the sound elements that are heard as part of the attended object (Woods and Colburn, 1992; Best *et al.*, 2007). For example, in one recent study, listeners did not show obligatory integration of spatial cues from simultaneously presented elements. Only when the simultaneous low-and high-frequency elements were perceived as part of the same object was integration observed. However, unlike in the current study, the elements that were perceived in different objects were not comprised of interleaved frequency components; instead, the low-and high-frequency elements were far removed from each other, spectrally. Nonetheless, manipulations that altered how strongly listeners grouped the low-and high-frequency elements together altered the amount of spatial-cue integration. If it were generally true that listeners only integrate spatial information across elements that are perceived as making up an object, then the perceived location of the complex should not be pulled by the target in the current experiment, since the target is not heard strongly as part of the complex. Thus, taken with past studies of how listeners group the current sound mixtures, we show here that integration of spatial cues contained in simultaneous elements (the target and the complex) is not predicted by whether the elements are heard as part of the same object. Instead, we find that across-frequency integration of spatial cues for the current simultaneous, spectrally interleaved elements is obligatory, regardless of whether or not the simultaneous elements are perceived as one object.

It is possible that subjects used in the current experiments actually heard the target as part of the complex, given that this was not explicitly measured here. However, this is unlikely based on the robustness of results from our past studies of how listeners group these kinds of sound mixtures (Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b). Another alternative is that there is no link between grouping and localization. However, this is inconsistent with previous observations showing that grouping influences what spatial cues listeners integrate when determining the perceived location of an object (Woods and Colburn, 1992; Hill and Darwin, 1996; Best *et al.*, 2007). Another possibility is that even a small contribution of an element to what an object sounds like can yield a large shift in where that object is perceived. This idea can be tested in future experiments by measuring whether the perceived location of simultaneous complex presented with an attenuated target is strongly pulled toward the location of a target even when it has little energy. However, preliminary tests in our laboratory suggest this is not the case (Schwartz and Shinn-Cunningham, 2008).

We believe that the most likely possibility is that listeners generally integrate spatial cues from only those sound elements that are part of an attended object. However, (1) listeners cannot perfectly filter out spatial cues in simultaneously presented components and (2) the degree to which they can filter out spatial cues in competing elements de-

pends on the frequency separation between within-object elements and interfering sound elements. In other words, just as the binaural system may be sluggish in processing spatial cues in the temporal dimension (Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998; Akeroyd and Summerfield, 1999), binaural analysis may also be coarse in frequency compared to monaural analysis (Holube *et al.*, 1998). One possible experiment that can test this conjecture would be to use similar stimuli but increase the fundamental frequency of the harmonic complex, so that the frequency separation of the harmonics is large enough that the spatial cues of the target are spectrally distinct from those of the complex. In this case, one could predict that the target will no longer influence the perceived location of the complex when the target is heard as part of the complex.

B. Pushing (repulsion)

There is no evidence of pushing in any of our single-object results. Instead, the target pulled both tone stream and complex in the absence of a competing object. However, when two objects were present in the scene and were perceived at different locations, they generally repelled one another. For instance, the perceived location of the complex was always displaced away from the tones location when tones were added to the mixture ($S_{45}\mathbf{C}_0T_0$ versus $\mathbf{C}_0T_0$ and $S_{45}\mathbf{C}_0T_{45}$ versus $\mathbf{C}_0T_{45}$). Similarly, the effect of moving the location of the tones in the two-object mixtures was to displace the perceived location of the complex in the opposite direction from the displacement of the tones ($S_0\mathbf{C}_0T_0$ versus $S_{45}\mathbf{C}_0T_0$ and $S_0\mathbf{C}_0T_{45}$ versus $S_{45}\mathbf{C}_0T_{45}$). Experiment 2 showed that across-object repulsion was often stronger when the stimulus rate was faster and weaker when the rate was slower, both when localizing the tones (in condition $\mathbf{S}_0C_0T_{45}$) and when localizing the harmonic complex (in conditions $S_0\mathbf{C}_0T_{45}$ and $S_{45}\mathbf{C}_0T_{45}$). This rate effect supports the idea that objects repel one another spatially, but that this effect decreases as the objects are more separated in time.

In the current study, only two objects were heard (the tone stream and the harmonic complex), and they were separated in time. One might, therefore, postulate that only temporally segregated objects repel one another. However, previous evidence for repulsion has been reported when objects overlap in time (Lorenzi *et al.*, 1999; Best *et al.*, 2005). Therefore, we suggest that repulsion occurs between objects (as opposed to within an object), and that this across-object repulsion weakens with temporal separation between the competing objects.

C. Assessing auditory segregation through pulling and pushing

If spatial information is integrated within an object, but objects repel one another, then spatial perception can be used to assess the perceptual segregation of elements comprising an auditory scene.

Kubovy and van Valkenburg (2001) argued that "perceptual boundaries" are important for the formation of auditory objects. In vision, perceptual boundaries, or edges, are determined by spatio-temporal discontinuities (Adelson and Bergen, 1991). In audition, spectro-temporal structure determines how objects form (Bregman, 1990; Darwin and Carlyon, 1995; Darwin, 1997; Van Valkenburg and Kubovy, 2003; Griffiths and Warren, 2004; Shinn-Cunningham, 2008). By parametrically varying the spectro-temporal features of sound in a mixture (e.g., in dimensions such as onset synchrony, harmonicity, common amplitude modulation, etc.), perceptual segregation can be manipulated, which should impact localization judgments. Specifically, if listeners judge sound elements to be coming from different locations, then the elements must belong to different perceptual objects. Conversely, if sound elements are heard as part of the same object, then listeners are likely to integrate the spatial information in the elements, leading to a pulling effect.

However, auditory objects are not always distinct (Rand, 1974; Liberman *et al.*, 1981; Moore *et al.*, 1986; Darwin, 1995; McAdams *et al.*, 1998; Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b; Shinn-Cunningham and Wang, 2008). These results are consistent with the fact that sound making up an auditory scene is transparent, with sounds from different sources adding together rather than obscuring each other (Bregman, 1990). Some studies suggest that it is necessary, but not sufficient, for sound elements to be perceived in distinct objects for them to be perceived as coming from distinct locations (Litovsky and Shinn-Cunningham, 2001; Best *et al.*, 2007). The current results (taken together with our past results investigating perceptual organization of these mixtures; Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b) demonstrate that integration of spatial cues can occur across elements that are not perceived within the same object. In other words, while spatial repulsion across elements may prove that the elements are heard in different objects, the current results show that integration of spatial cues across element does *not* prove that the elements are heard in the *same* object. Still, using a continuous measure such as perceived location can provide a bound on when different objects are perceived as distinct, even though it cannot rule out cases when two objects are heard, but are perceived at the same location. Future work can assess the degree to which spatial measures of across-element spatial integration and across-object spatial repulsion give insights into auditory scene analysis.

V. CONCLUSIONS

These data show that there is repulsion between the perceived locations of auditory objects, and that this repulsion tends to decrease with increasing temporal separation of the objects. Moreover, spatial cues in one sound element are often integrated with other spatial cues, pulling the perceived location of an attended object toward the location of the individual element, both for single-and two-object sound mixtures. We observed some weak integration of spatial cues across time, consistent with binaural sluggishness. However, this across-time pulling was only present in some single-object mixtures and was not observed for any of the two-object mixtures used here. We found evidence for an obligatory integration of the spatial cues in a simultaneous target element with those of a spectrally interleaved harmonic com-

plex, an effect that was independent of repetition rate. Taken together with our companion grouping experiments using the same sets of stimuli (Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b), and in contrast with previous results (Woods and Colburn, 1992; Hill and Darwin, 1996; Best *et al.*, 2007), the current results show that spatial cues in elements that do not contribute strongly to the perceived content of an auditory object can nonetheless strongly pull the perceived location of that object. We suggest that spatial repulsion is an effect observed between objects, while spatial cue integration is an effect observed either within an object or across sound elements whose spatial cues cannot be resolved in spatial computations.

ACKNOWLEDGMENTS

This work was supported by grant from the National Institutes of Health (DC05778-02 and DC009477) to B.G.S.-C. A.D.-P. would like to acknowledge the Bogue Research Fellowship, UCL. Sigrid Nasser helped with subject recruitment and data collection.

[1]Based on pilot data, we selected the allowable range of ILD matches to have a maximum magnitude of 20 dB. While this range was sufficiently large to guarantee that our pilot subjects never reached the maximum allowable value, this was not the case for all of the subjects tested in the formal experiment. In general, we used simple, paired *t*-tests to check for the statistical significance of effects of interest. However, *t*-tests assume Gaussian-distributed matches. Thus, for some comparisons involving matches to objects perceived to the side, the response distributions of the ILD matches might be skewed due to the response limitations, violating the assumptions of a parametric *t*-test. Therefore, we performed nonparametric Wilcoxon signed-rank tests on comparisons involving the perceived locations of objects originating from the side. These comparisons are denoted with italics in Table I. Taking a fairly conservative approach, we further excluded comparisons between pairs of conditions when more than a quarter of the matches in each of the conditions to be compared had magnitudes equal to or greater than 18 dB (within 2 dB of the maximum allowable response) to prevent over-interpreting the results. Comparisons that were excluded due to responses near the allowable maximum are indicated by ellipses in the tables summarizing the statistical comparisons.

[2]We hypothesized that the matched location would be further away from midline for the faster repetition rate (with no underlying distribution assumed for the amount of repulsion). Therefore, when testing the significance of rate on repulsion in these two-object conditions, we used one-tailed Wilcoxon signed-rank tests.

Adelson, E. H., and Bergen, J. R. (**1991**). *The Plenoptic Function and the Elements of Early Vision* (MIT, Cambridge, MA).

Akeroyd, M. A., and Summerfield, A. Q. (**1999**). "A binaural analog of gap detection," J. Acoust. Soc. Am. **105**, 2807–2820.

Anstis, S., and Saida, S. (**1985**). "Adaptation to auditory streaming of frequency-modulated tones," J. Exp. Psychol. Hum. Percept. Perform. **11**, 257–271.

Arbogast, T. L., Mason, C. R., and Kidd, G. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Bernstein, L. R., and Trahiotis, C. (**1985**). "Lateralization of sinusoidally amplitude-modulated tones: effects of spectral locus and temporal variation," J. Acoust. Soc. Am. **78**, 514–523.

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

Best, V., van Schaik, A., Jin, C., and Carlile, S. (**2005**). "Auditory spatial perception with sources overlapping in frequency and time," Acta. Acust. Acust. **91**, 421–428.

Best, V., Gallun, F. J., Carlile, S., and Shinn-Cunningham, B. G. (**2007**). "Binaural interference and auditory grouping," J. Acoust. Soc. Am. **121**, 1070–1076.

Braasch, J., and Hartung, K. (**2002**). "Localization in the presence of a distracter and reverberation in the frontal horizontal plane. I. Psychoacoustical data," Acta. Acust. Acust. **88**, 942–955.

Bregman, A. S. (**1978**). "Auditory streaming is cumulative," J. Exp. Psychol. Hum. Percept. Perform. **4**, 380–387.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, Mass.).

Buell, T. N., Trahiotis, C., and Bernstein, L. R. (**1991**). "Lateralization of low-frequency tones—Relative potency of gating and ongoing interaural delays," J. Acoust. Soc. Am. **90**, 3077–3085.

Butler, R. A., and Naunton, R. F. (**1964**). "Role of stimulus frequency and duration in the phenomenon of localization shifts," J. Acoust. Soc. Am. **36**, 917–922.

Carlyon, R. P., Cusack, R., Foxton, J. M., and Robertson, I. H. (**2001**). "Effects of attention and unilateral neglect on auditory stream segregation," J. Exp. Psychol. Hum. Percept. Perform. **27**, 115–127.

Cherry, E. C. (**1953**). "Some experiments on the recognition of speech, with one and with two ears," J. Acoust. Soc. Am. **25**, 975–979.

Culling, J. F., and Colburn, H. S. (**2000**). "Binaural sluggishness in the perception of tone sequences and speech in noise," J. Acoust. Soc. Am. **107**, 517–527.

Culling, J. F., and Summerfield, Q. (**1998**). "Measurements of the binaural temporal window using a detection task," J. Acoust. Soc. Am. **103**, 3540–3553.

Cusack, R., Deeks, J., Aikman, G., and Carlyon, R. P. (**2004**). "Effects of location, frequency region, and time course of selective attention on auditory scene analysis," J. Exp. Psychol. Hum. Percept. Perform. **30**, 643–656.

Darwin, C. J. (**1995**). "Perceiving vowels in the presence of another sound: a quantitative test of the "Old-plus-New" heuristic," in *Levels in Speech Communication: Relations and Interactions: A Tribute to Max Wajskop*, edited by C. Sorin, J. Mariani, and H. Meloni (Elsevier, Amsterdam), pp. 1–12.

Darwin, C. J. (**1997**). "Auditory grouping," Trends Cogn. Sci. **1**, 327–333.

Darwin, C. J., and Carlyon, R. P. (**1995**). in *Hearing*, edited by B. C. J. Moore (Academic, San Diego, CA), p. 387.

Darwin, C. J., and Hukin, R. W. (**1997**). "Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity," J. Acoust. Soc. Am. **102**, 2316–2324.

Darwin, C. J., and Hukin, R. W. (**1999**). "Auditory objects of attention: The role of interaural time differences," J. Exp. Psychol. Hum. Percept. Perform. **25**, 617–629.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.

Gardner, M. B. (**1969**). "Image fusion, broadening, and displacement in sound location," J. Acoust. Soc. Am. **46**, 339–349.

Good, M. D., and Gilkey, R. H. (**1996**). "Sound localization in noise: The effect of signal-to-noise ratio," J. Acoust. Soc. Am. **99**, 1108–1117.

Grantham, D. W., and Wightman, F. L. (**1978**). "Detectability of varying interaural temporal differences," J. Acoust. Soc. Am. **63**, 511–523.

Griffiths, T. D., and Warren, J. D. (**2004**). "What is an auditory object?," Nat. Rev. Neurosci. **5**, 887–892.

Heller, L. M., and Trahiotis, C. (**1996**). "Extents of laterality and binaural interference effects," J. Acoust. Soc. Am. **99**, 3632–3637.

Hill, N. I., and Darwin, C. J. (**1996**). "Lateralization of a perturbed harmonic: effects of onset asynchrony and mistuning," J. Acoust. Soc. Am. **100**, 2352–2364.

Holube, I., Kinkel, M., and Kollmeier, B. (**1998**). "Binaural and monaural auditory filter bandwidths and time constants in probe tone detection experiments," J. Acoust. Soc. Am. **104**, 2412–2425.

Kollmeier, B., and Gilkey, R. H. (**1990**). "Binaural forward and backward-masking-evidence for sluggishness in binaural detection," J. Acoust. Soc. Am. **87**, 1709–1719.

Kubovy, M., and Van Valkenburg, D. (**2001**). "Auditory and visual objects," Cognition **80**, 97–126.

Lee, A. K. C., Babcock, S., and Shinn-Cunningham, B. G. (**2008**). "Measuring the perceived content of auditory objects using a matching paradigm," J. Assoc. Res. Otolaryngol. **9**, 388–397.

Lee, A. K. C., and Shinn-Cunningham, B. G. (**2008a**). "Effects of frequency disparities on trading of an ambiguous tone between two competing auditory objects," J. Acoust. Soc. Am. **123**, 4340–4351.

Lee, A. K. C., and Shinn-Cunningham, B. G. (**2008b**). "Effects of reverberant spatial cues on attention-dependent object formation," J. Assoc. Res.

Otolaryngol. **9**, 150–160.
Liberman, A. M., Isenberg, D., and Rakerd, B. (**1981**). "Duplex perception of cues for stop consonants—Evidence for a phonetic mode," Percept. Psychophys. **30**, 133–143.
Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.
Litovsky, R. Y., and Shinn-Cunningham, B. G. (**2001**). "Investigation of the relationship among three common measures of precedence: Fusion, localization dominance, and discrimination suppression," J. Acoust. Soc. Am. **109**, 346–358.
Lorenzi, C., Gatehouse, S., and Lever, C. (**1999**). "Sound localization in noise in normal-hearing listeners," J. Acoust. Soc. Am. **105**, 1810–1820.
McAdams, S., Botte, M. C., and Drake, C. (**1998**). "Auditory continuity and loudness computation," J. Acoust. Soc. Am. **103**, 1580–1591.
McFadden, D., and Pasanen, E. G. (**1976**). "Lateralization at high frequencies based on interaural time differences," J. Acoust. Soc. Am. **59**, 634–639.
Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1986**). "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," J. Acoust. Soc. Am. **80**, 479–483.
Rand, T. C. (**1974**). "Dichotic release from masking for speech," J. Acoust. Soc. Am. **55**, 678–680.
Schwartz, A., and Shinn-Cunningham, B. G. (**2008**). "The influence of ambiguous grouping cues on an auditory object's perceived spectral content and location," in Mid-Winter Meeting of the Association for Research in Otolaryngology, Phoenix, AZ.
Shinn-Cunningham, B. G. (**2008**). "Object-based auditory and visual attention," Trends Cogn. Sci. **12**, 182–186.
Shinn-Cunningham, B. G., Ihlefeld, A., Satyavarta, and Larson, E. (**2005**). "Bottom-up and top-down influences on spatial unmasking," Acta. Acust. Acust. **91**, 967–979.
Shinn-Cunningham, B. G., Kopco, N., and Martin, T. J. (**2005b**). "Localizing nearby sound sources in a classroom: Binaural room impulse response," J. Acoust. Soc. Am. **117**, 3100–3115.
Shinn-Cunningham, B. G., Lee, A. K. C., and Oxenham, A. J. (**2007**). "A sound element gets lost in perceptual competition," Proc. Natl. Acad. Sci. U.S.A. **104**, 12223–12227.
Shinn-Cunningham, B. G., and Wang, D. (**2008**). "Influences of auditory object formation on phonemic restoration," J. Acoust. Soc. Am. **121**, 295–301.
Trahiotis, C., and Stern, R. M. (**1989**). "Lateralization of bands of noise—Effects of bandwidth and differences of interaural time and phase," J. Acoust. Soc. Am. **86**, 1285–1293.
Van Valkenburg, D., and Kubovy, M. (**2003**). "In defense of the theory of indispensable attributes," Cognition **87**, 225–233.
Woods, W. S., and Colburn, H. S. (**1992**). "Test of a model of auditory object formation using intensity and interaural time difference discrimination," J. Acoust. Soc. Am. **91**, 2894–2902.

Effects of source-to-listener distance and masking on perception of cochlear implant processed speech in reverberant rooms

Nathaniel A. Whitmal III[a] and Sarah F. Poissant
Department of Communication Disorders, University of Massachusetts, Amherst, Massachusetts 01003

(Received 22 October 2008; revised 29 July 2009; accepted 5 August 2009)

Two experiments examined the effects of source-to-listener distance (SLD) on sentence recognition in simulations of cochlear implant usage in noisy, reverberant rooms. Experiment 1 tested sentence recognition for three locations in the reverberant field of a small classroom (volume=79.2 m^3). Subjects listened to sentences mixed with speech-spectrum noise that were processed with simulated reverberation followed by either vocoding (6, 12, or 24 spectral channels) or no further processing. Results indicated that changes in SLD within a small room produced only minor changes in recognition performance, a finding likely related to the listener remaining in the reverberant field. Experiment 2 tested sentence recognition for a simulated six-channel implant in a larger classroom (volume=175.9 m^3) with varying levels of reverberation that could place the three listening locations in either the direct or reverberant field of the room. Results indicated that reducing SLD did improve performance, particularly when direct sound dominated the signal, but did not completely eliminate the effects of reverberation. Scores for both experiments were predicted accurately from speech transmission index values that modeled the effects of SLD, reverberation, and noise in terms of their effects on modulations of the speech envelope. Such models may prove to be a useful predictive tool for evaluating the quality of listening environments for cochlear implant users. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3216912]

PACS number(s): 43.66.Ts, 43.71.Es, 43.71.Ky [RYL] Pages: 2556–2569

I. INTRODUCTION

The intelligibility of speech in rooms is affected by reverberation. Reverberant sound energy typically creates a temporal "smearing" of speech that imposes overlap masking on contiguous phonemes, lengthens the durations of words, and fills quiet and/or low-intensity speech segments with unwanted sound (Bolt and MacDonald, 1949; Houtgast and Steeneken, 1985; Nabelek *et al.*, 1989; Dreschler and Leeuw, 1990; Helfer, 1994; Culling *et al.*, 2003). As a result, intelligibility decreases in conjunction with the reductions in speech envelope modulation depth imposed by temporal smearing (Houtgast and Steeneken, 1985). Competing speech and other ambient noises, being similarly affected, interact with the distorted speech to reduce intelligibility more than either noise or reverberation would alone (Duquesnoy and Plomp, 1980; Nabelek and Robinson, 1982; Crandell and Smaldino, 2000). These reductions are particularly severe for listeners with impaired hearing, who typically require less noise and reverberation to achieve the same intelligibility as listeners with normal hearing (Finitzo-Hieber and Tillman, 1978; Duquesnoy and Plomp, 1980; Helfer and Wilber, 1990).

The temporal effects of reverberation on speech may pose a particular challenge for cochlear implant (CI) users, who receive their auditory cues from temporal envelope modulations in a limited number of spectral channels. Temporal envelope modulations suffice to provide intelligible speech for as few as four spectral channels under ideal conditions (Shannon *et al.*, 1995; Dorman *et al.*, 1997). However, when the channels' envelope modulations are low-pass filtered with cutoff frequencies in the 8–16 Hz range, intelligibility decreases (Fu and Shannon, 2000; Xu and Zheng, 2007). Reverberation, which can act as a low-pass filter for envelope modulations in this frequency range (Houtgast and Steeneken, 1985), has similarly been shown to degrade intelligibility for listening through actual or simulated implants in rooms that listeners with normal hearing would find acceptable (Iglehart, 2004; Poissant *et al.*, 2006).

Studies of intelligibility for implant users in reverberant spaces have typically focused on the use of frequency modulation (FM) or sound field devices to improve intelligibility (Crandell *et al.*, 1998; Iglehart, 2004; Anderson *et al.*, 2005), with mixed results. Few studies have focused on the specific effects of reverberation and noise on implant processed speech. Poissant *et al.* (2006) used simulations of both reverberation (Allen and Berkley, 1979; Peterson, 1986) and implant processing (Qin and Oxenham, 2003) to investigate the effects of reverberation on the intelligibility of implant processed sentence key words in a small classroom. Results showed that speech recognition scores decreased in conjunction with decreases in the number of spectral channels and/or the room's uniform absorption coefficient (α). For example, intelligibility scores for a six-channel implant simulation in quiet decreased from 87% correct to 22% correct when the reverberation time (RT_{60}) was increased from 0 to 520 ms, a RT_{60} value considered acceptable by ANSI classroom standards (ANSI, 2002). These intelligibility decreases were subsequently worsened by mixing the target speech with either speech-spectrum noise or two-talker babble, both of which further reduced intelligibility for speech in quiet with RT_{60}

[a]Author to whom correspondence should be addressed. Electronic mail: nwhitmal@comdis.umass.edu

=266 ms by nearly 40% when added at a +8 dB signal-to-noise ratio (SNR). No significant interaction between levels of SNR and levels of RT_{60} or masker type were observed, a finding that contrasted markedly with previous data showing more severe effects for noise and reverberation combined than for the combination of their individual effects (Nabelek and Mason, 1981; Loven and Collins, 1988; Helfer and Wilber, 1990; Payton *et al.*, 1994) or more efficient masking for competing speech than for speech-spectrum noise in both simulated and actual implant systems (Qin and Oxenham, 2003; Stickney *et al.*, 2004). Statistically significant positive correlations between intelligibility scores and computed speech transmission index (STI) values (Houtgast and Steeneken, 1985) for each noise type indicated a strong relationship between intelligibility and the envelope modulations of the vocoded reverberated speech. The STI data of Poissant *et al.* (2006) used a modified computation intended for use with nonlinear signal processors such as those common to CIs (Goldsworthy and Greenberg, 2004).

The listening difficulties observed by Poissant *et al.* (2006) are not unexpected, given the placement of the listener in the room's reverberant field. However, the degree of difficulty observed was remarkable considering the small size of the room (volume=79.2 m^3) and short source-to-listener distance (SLD) of 4 m, and was far greater than the difficulties normally associated with the magnitude of STI values computed in that study. Under the study's conditions, most of the sound energy reaching the listener consisted of "early" reflections arriving 50–80 ms after the direct sound. The auditory system usually integrates early reflections together with the direct sound to increase both the perceived loudness and the intelligibility of speech (Lochner and Burger, 1964; Latham, 1979; Bradley, 1986a, 1986b). Such increases can be particularly important in noisy conditions (Bradley *et al.*, 2003). Studies exploring the relationship between RT_{60} and intelligibility in both quiet and noisy classrooms (Bistafa and Bradley, 2000; Yang and Hodgson, 2006) indicate that the RT_{60} value giving the best intelligibility in noise increases as room volume and noise-to-signal ratio increase. These increases in RT_{60} were associated with better intelligibility for hearing impaired subjects when sources of noise (e.g., neighboring students) were closer to the listener than sources of speech (e.g., instructors and/or audio equipment). There, early reflections comprised a greater proportion of speech energy than of noise energy, compensating in part for the proximity of the noise and enhancing the intelligibility of the speech.

At present, it is not known how well implant users can integrate early reflections with direct sound as described above. The results of Poissant *et al.* (2006) suggest that signals comprised largely of early reflections may be less intelligible for implant users than signals comprised largely of direct arrivals. This hypothesis, if true, has important implications for implant users since recommendations for improving classroom acoustics often include increasing the level of early reflections (Siebein *et al.*, 2000; Bradley *et al.*, 2003). One simple way to evaluate this hypothesis is to compare intelligibility scores for speech received at small SLDs (having strong contributions from direct sound) with those for speech received at large SLDs (having strong contributions from reflected sound). Listeners with normal temporal integration abilities would be expected to recognize speech at the front and rear of the room with equal facility [assuming equal sound pressure levels (SPLs)]; listeners showing deficits in integration ability would be expected to show significant differences between the two locations. This approach is followed in the present work.

Another remarkable aspect of the Poissant *et al.* (2006) study was the strong correlation observed between computed STI values and measured intelligibility scores. Although the standard STI computation is widely used in assessment of listening rooms and sound reinforcement systems, it is not recommended for use in assessing vocoder systems or predicting intelligibility for hearing impaired listeners (IEC, 2002). Goldsworthy and Greenberg (2004) noted that nonlinear operations commonly found in speech processing systems (e.g., power spectrum subtraction) produce artifacts that distort the association between STI values and intelligibility scores. They subsequently proposed several modifications to the STI that appeared to eliminate these distortions. While they noted that the STI would be a good candidate measure for predicting intelligibility for CI users, they did not present STI data for either simulated or actual CI processed speech. The data of Poissant *et al.* (2006) for simulated CI speech with 6-channel and 12-channel vocoders supported their assumptions while drawing attention to effects of subject proficiency. Specifically, the relationship between STI values and intelligibility scores was shown to depend on the number of vocoder channels, with individual STI values mapping to higher scores for the 12-channel vocoder than for the 6-channel vocoder. This finding reflects a fundamental difference between the STI, which does not account for subject proficiency, and measures like the Articulation Index (ANSI, 1969) and Speech Intelligibility Index (ANSI, 1997) that can model subject proficiency (reflected through hearing thresholds) for intelligibility prediction with individual subjects. The influence of proficiency on STI-based predictions has been addressed in work with subjects with sensorineural hearing losses (HLs) (Dreschler and Leeuw, 1990; Duquesnoy and Plomp, 1980), but not in work with implant users. The present work directly examines the relationship between proficiency, STI, and intelligibility for implant simulations, where proficiency is modeled as the number of vocoder channels available to the listener.

The purpose of the present study was to investigate the effects of SLD on the intelligibility of CI processed speech and on the STI as a predictor of CI speech intelligibility in both quiet and noisy reverberant rooms. Two experiments were conducted. Experiment 1 investigated the effects of SLD and number of available spectral channels on intelligibility in quiet and in noise for implant processed speech in the small classroom evaluated by Poissant *et al.* (2006) to determine whether reducing SLD would lead to better performance. The results of Experiment 1 provide a measure of the listener's ability to take advantage of early reflections in that small room. Experiment 2 investigated the individual and combined effects of reverberation, SLD, and noise on processed speech intelligibility in a second, slightly larger

classroom. Reverberation in this room was determined by specifying various values of α; this, in combination with the larger room volume, makes it possible to evaluate listener performance at SLDs that are both less than and greater than the room's critical distance (i.e., the SLD at which direct and reverberant sound energies are equal).

The STI analysis in the present work uses the traditional STI approach (Houtgast and Steeneken, 1985), which differs from the newer envelope-regression based STI approach of Goldsworthy and Greenberg (2004) chosen by Poissant *et al.* (2006) for its ability to accommodate nonlinearly processed signals. The Goldsworthy/Greenberg (2004) approach produces STI values that correlate well with traditional STI values for unvocoded speech in noisy and/or reverberant conditions. For vocoded speech, Poissant *et al.* (2006) found that the Goldsworthy/Greenberg (2004) approach produced STI values that were compressed nonlinearly into a narrower range than traditionally produced values. The upper bound of this range was dependent (in an undetermined manner) on the number of vocoder channels used. These unexamined phenomena are a reflection of the vocoder's effects on the speech signal, and are worthy of study in a separate investigation that focuses on the mathematics of vocoder processing and STI computation. To expedite the present work, we chose to use the traditional STI approach, which, in addition to being widely studied, is currently the only STI version that has been shown to correlate consistently with measures of early reflection benefit (Bradley *et al.*, 1999, 2003).

II. EXPERIMENT 1: EFFECTS OF DISTANCE AND NUMBER OF SPECTRAL CHANNELS ON INTELLIGIBILITY OF PROCESSED SENTENCES

A. Methods

1. Subjects

Twelve adult listeners (ten females and two males) participated in experiment 1. The subjects' ages ranged from 19 to 31 years (mean age=23.8 years). All subjects were native speakers of American English who had passed a screening for normal hearing (thresholds $\leq$20 dB HL). None of the subjects had participated in previous simulation experiments. The subjects were compensated for their participation with partial course credit.

2. Materials

Stimuli for experiment 1 were the same as those used by Poissant *et al.* (2006), and are described briefly here. The stimuli consisted of 360 sentences (Helfer and Freyman, 2004), each containing three key words in common use (Francis and Kucera, 1982). The sentences were assigned to 1 of 24 topics (e.g., food, clothing, and politics) used to help listeners direct their attention to the target speaker when sentences were heard in the presence of competing speakers. The sentences were uttered by a female speaker with an American English dialect and digitally recorded in a sound-treated booth (IAC 1604) with 16-bit resolution at a 22 050 Hz sampling rate.

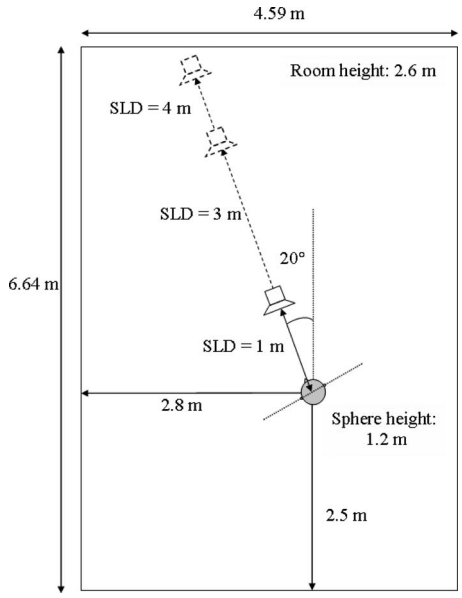


FIG. 1. Schematic for the reverberation simulation of experiment 1, illustrating the dimensions of the room, the orientation of the sphere modeling the listener's head, and the relative positions of the sphere and the target source within the room (adapted from Poissant *et al.*, 2006).

3. Signal processing

The signal processing utilized in this investigation was used previously in Poissant *et al.*, 2006 and consists of (optional) noise addition followed in sequence by reverberation simulation and (optional) CI simulation. These processes are described below.

Noise addition. The sentence recordings described above were input to subsequent simulators either as recorded in quiet or with speech-spectrum noise added at SNRs of +8 or +18 dB, with one exception: SNRs for "unprocessed" conditions (described below) included −8 and +8 dB. The +8 and +18 dB SNRs were chosen to facilitate direct comparisons of the present results with data from previous studies using the same stimuli and simulators (Poissant *et al.*, 2006; Whitmal *et al.*, 2007); the −8 dB SNR replicated a condition used in previous studies with the topic sentence recordings (Helfer and Freyman, 2004, 2005). Maskers for each sentence were derived from scaled segments of the speech-spectrum noise as in Poissant *et al.*, 2006.

Reverberation simulation. The reverberation simulation utilized an image-source software model (Allen and Berkley, 1979; Peterson, 1986) of a small rectangular classroom with uniform, frequency-independent absorption on all surfaces. The dimensions for the ideal classroom were taken from a real rectangular classroom at the University of Massachusetts Amherst. Figure 1 provides details of the dimensions of the room, the orientation of a simulated listener's head (modeled as a sphere), and the relative positions of the listener's head and the target source. The source (modeled by the software as omnidirectional) was placed at each of three locations located 1, 3, or 4 m from the listener at a 0° azimuth. Since most real-world sources exhibit some directionality that can improve intelligibility by boosting direct-to-reverberant energy ratio, the model used here represents a worst-case scenario. α was set to 0.25, a previously explored value (Pois-

TABLE I. Center frequencies and bandwidths, expressed in hertz, for each of three experimental vocoder systems.

Channels	Parameters	Band values											
6	CF	180	446	885	1609	2803	4773						
	BW	201	331	546	901	1487	2453						
12	CF	124	224	353	519	731	1005	1355	1806	2385	3128	4084	5310
	BW	88	113	145	186	239	307	395	507	651	836	1074	1379
24													
Bands 1–12	CF	101	145	194	251	315	387	469	562	668	787	923	1077
	BW	41	47	53	60	68	77	87	99	112	127	144	163
Bands 13–24	CF	1251	1448	1671	1925	2212	2538	2906	3324	3798	4335	4944	5634
	BW	185	210	238	269	305	346	392	444	503	571	647	733

CF=center frequency of channel and BW=bandwidth of channel.

sant *et al.*, 2006) expected to provide challenging listening conditions. The theoretical critical distance (Kuttruff, 1979) for this room configuration is

$$d_c = \frac{1}{4}\sqrt{\frac{S\alpha}{\pi}} = 0.774 \text{ m}, \quad (1)$$

with S being the total surface area of the room's walls, floor, and ceiling. All listening positions would therefore be expected to be in the reverberant field. Accordingly, direct-to-reverberant energy ratios of −2.9, −12.4, and −14.5 dB were measured at SLDs of 1, 3, and 4 m, respectively.

The room models were used to generate impulse responses for each SLD condition. Each impulse response was convolved with the sentence recordings to produce reverberant speech. For sentences in noise, the sequence of noise addition and convolution had the effect of placing the speech and noise in the same source location. RT_{60} values derived from reversed-time integrations of the three squared impulse responses (Schroeder, 1965) in the 1000 Hz octave-band over the −5 to −30 dB decay range were approximately 520 ms, well within the recommended range for classrooms of this size (ANSI, 2002). It should be noted that the RT_{60} value for this room is larger than the RT_{60} of 425 ms reported by Poissant *et al.* (2006) for the same room. There, RT_{60} was predicted from Sabine's (1922) theoretical formula as

$$RT_{60} = \frac{0.161V}{S\alpha}, \quad (2)$$

where V was the room volume in m^3. The discrepancy between predicted and measured RT_{60} values is consistent with the recent work of Lehmann and Johansson (2008), who showed that the Sabine (1922) formula tends to underpredict RT_{60} values in image-source room simulations. RT_{60} values used in the remainder of the paper will therefore refer only to times derived from impulse responses using Schroeder's (1965) method.

The effect of reverberation on presentation level was assessed by comparing the A-weighted rms level of a 15-s recording of anechoic speech (i.e., measured after setting $\alpha = 1$ within the simulation) with the A-weighted level of the same speech measured at each of the three listener positions. For anechoic speech of 65 dBA at 1 m, measured speech levels at the 1, 3, and 4 m positions were 67.7, 64.8, and 63.5 dBA, respectively. These small differences in level as a function of SLD are consistent with theoretical predictions for a classroom of this size (Barron and Lee, 1988; Sato and Bradley, 2008). All speech signals were subsequently presented to the subjects at 65 dBA, with the level held constant to differentiate the effects of reflection patterns at each position from any effects of level difference.

Implant processing simulation. Reverberated sentences were either processed by one of three tone-excited channel vocoder systems (i.e., systems with 6, 12, and 24 channels) implemented in MATLAB (Mathworks, Natick, MA), or subjected to no further processing or reduction in bandwidth (subsequently referred to as unprocessed). The implementations of the vocoder systems followed those of Qin and Oxenham (2003). For each vocoder, the input speech was filtered into contiguous frequency bands in the 80–6000 Hz range, each with equal width on an equivalent-rectangular-bandwidth scale (Glasberg and Moore, 1990). Center frequencies and bandwidths for each of the processors are provided in Table I. The envelope of each band was extracted via half-wave rectification and low-pass filtering and used to modulate a pure tone located at the band's center frequency. The bandwidth of the low-pass filter was the smaller of 300 Hz or half the analysis bandwidth. The tones for all bands were then scaled and added electronically to produce a simulated implant processed signal with a rms level of 65 dBA.

4. Procedure

Subjects listened to the 360 sentences while seated in a double-walled sound-treated booth (IAC 1604) during one 90-min listening session. Subjects were given breaks in the middle of each session. Each combination of the 3 noise conditions, 4 processing conditions, and 3 SLDs was used to process 1 of 36 ten-sentence lists. The presentation order of conditions for the subjects was determined by a 36×12 Latin rectangle, with the ordering of sentences for each list randomized. The presentation level for the sentences (65 dBA) was calibrated daily using repeated loops of the speech-spectrum noise described above.

Custom MATLAB software (executed on a laptop computer inside the test booth) was used to present the sentences

to the subject and to score the number of key words correctly recognized by her or him. The laptop screen prompted the subject with the word "Ready?" and the sentence topic exactly 2 s before the sentence was presented. The sentence was then retrieved from the remote computer's hard disk, converted to an analog signal by the computer's sound card (SigmaTel High Definition Audio Codec) using 16-bit resolution at a 22 050 Hz sampling rate, and input to a headphone amplifier (Behringer Pro-XL HA4700) driving a pair of Sennheiser HD580 circumaural headphones. The subject then typed the sentence as heard into a text window and submitted it to the software. Subjects were instructed to type any portion of the sentence that was intelligible, or "I don't know" if the sentence was completely unintelligible. The subjects' typed responses were later proofread by the authors, with obvious spelling mistakes and homophone substitutions corrected prior to scoring.

Practice materials were limited to ten sentences per vocoder, presented without feedback at the beginning of the experiment. The ten sentences were processed with the 3 m SLD simulation, with five presented in quiet and five presented in speech-spectrum noise at +8 dB SNR. The sentences used for practice were not used in the main experiment.

B. Computation of STI values

The STI is a frequency-weighted average of seven octave-band apparent signal-to-noise ratios (aSNRs), given as

$$\mathrm{STI} = \frac{\Sigma_{i=1}^{7} w_i(\mathrm{aSNR}_i) + 15}{30}, \tag{3}$$

where w_i was an empirically derived weight for band i (Houtgast and Steeneken 1985). aSNR values can range from -15 (representing poor intelligibility) to +15 dB (representing excellent intelligibility); consequently, STI values range from 0 (poor intelligibility) to 1 (excellent intelligibility). aSNR values are calculated from modulation transfer functions (MTFs) that quantify changes in modulation depth

$$\mathrm{aSNR}_i = 10 \log_{10}\left[\frac{\mathrm{MTF}_i}{1 - \mathrm{MTF}_i}\right], \tag{4}$$

where $i = 1, 2, \ldots, 7$ denotes the octave band and MTF_i denotes measurable reductions in modulation depth for band i, measured in and averaged over 14 one-third-octave spaced modulation frequencies between 0.63 and 12.5 Hz. For the case of speech-in-noise in an ideal room with a diffuse reverberant field, the theoretical modulation depth reduction $m_i(f)$ in band i at modulation frequency f is

$$m_i(f) = \frac{1}{\sqrt{1 + (2\pi f\ \mathrm{RT}_{60}/13.8)^2}}\left[\frac{1}{1 + 10^{-\mathrm{SNR}_i/10}}\right], \tag{5}$$

where SNR_i was the SNR in decibels for band i (Houtgast *et al.*, 1980). Equation (5) illustrates two factors affecting envelope modulations: low-pass filtering attributable to reverberation, and frequency-independent attenuation attributable to additive noise. In practice, MTF values are derived from responses to input probe signals consisting of amplitude modulated noise or speech (Steeneken and Houtgast, 1982; Payton and Braida, 1999).

In the present work, STI values for all conditions of reverberation and noise were measured in MATLAB on a Pentium 4 personal computer, using the approach of Houtgast and Steeneken (1973). Briefly, MTFs were computed from probe signals consisting of speech-spectrum noise with 100% sinusoidal intensity modulation at each of the 14 modulation frequencies mentioned above. The probe signals were convolved with the room impulse responses and filtered (using eighth-order Butterworth filters) into one of the seven octave bands. Intensity envelopes for the band-limited signals were then computed by squaring and low-pass filtering the signals with a 300 Hz fourth-order Butterworth low-pass filter. The modulation depths of the intensity envelopes for each frequency were then measured and averaged across frequency to produce a modulation index.

C. Results

1. Intelligibility scores

Intelligibility scores for experiment 1 were derived from the percentage of correctly repeated key words per condition. Mean intelligibility scores for each channel configuration are shown in Fig. 2. As expected, the best performance was observed for unprocessed speech in quiet, with average scores ranging between 92.8% and 94.2% correct. Average scores in quiet for the 24-channel vocoder at SLDs of 1 and 3 m were within the same range as the unprocessed scores: average scores for the 12- and 6-channel vocoders were considerably lower (76.5% and 32.3% correct, respectively). This relationship between available channels and intelligibility scores is similar to that observed by Poissant *et al.* (2006). The best performance in quiet for each vocoder was observed at the 3 m SLD. Scores in quiet at the 1 m SLD were slightly lower than 3 m SLD scores, with average differences of 0.8%, 6.1%, and 5.0% observed for 24-, 12-, and 6-channel vocoders, respectively. Scores in quiet for each vocoder at the 4 m SLD were (on average) 7.8% below corresponding scores at the 1 m SLD.

Performance in all listening conditions decreased considerably in the presence of noise. Adding noise at a +18 dB SNR reduced average scores for 24-, 12-, and 6-channel vocoders by 3.1%, 7.4%, and 9.3%, respectively. In each case, the largest reductions were observed for scores at the 3 m SLD, which subsequently became less than or equal to scores at 1 m. Adding noise at a +8 dB SNR reduced average scores overall by 22.8%, 35.1%, and 22.6%, respectively. Effects of decreasing SLD were most evident at the +8 dB SNR. Scores decreased by between 14% and 20% for 24- and 12-channel vocoders when SLD was increased from 3 to 4 m, while a smaller decrease of 6% (presumably denoting a floor effect) was observed for the 6-channel vocoder. In contrast, unprocessed scores in noise at +8 dB SNR dropped (on average) only 4.5% below corresponding scores in quiet, and showed little variation with respect to SLD. Scores for unprocessed speech at -8 dB SNR (a challenging SNR) were less than 3% correct for all SLDs.

Subject scores were converted to rationalized arcsine

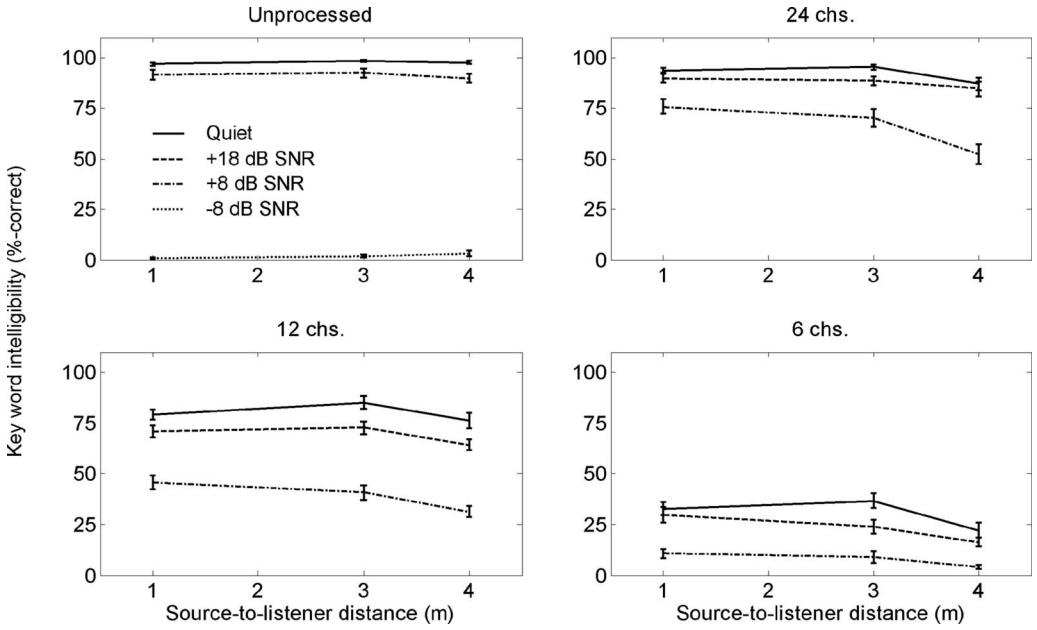


FIG. 2. Speech recognition performance for unprocessed/natural speech and 24-, 12-, and 6-channel vocoded speech in quiet and in noise as a function of SLD in the experiment 1 room simulation. Error bars represent ±1 standard error.

units (Studebaker, 1985) and input to a repeated-measures analysis of variance of intelligibility scores. Within-subject factors for the analysis of variance included the number of channels ($F[3,306]=1188.91$, $p<0.0001$), SLD ($F[2,306]=30.67$, $p<0.0001$), and SNR ($F[3,306]=687.11$, $p<0.0001$), all of which were statistically significant. All first-order interactions between main factors were significant. *Post hoc* tests using the Tukey honestly significant difference criterion ($\alpha=0.05$) indicated that (a) scores at 4 m were significantly lower than scores at 1 or 3 m, (b) scores at 1 and 3 m were not significantly different from each other, (c) scores for 6-, 12-, and 24-channel processors were all significantly different from each other, and (d) scores for each SNR were all significantly different from each other.

2. STI values

STI values for the listening conditions of experiment 1 are shown in Fig. 3. Measured values are represented by filled symbols; theoretical values in the reverberant field [computed according to Eq. (5)] are represented by unfilled symbols. The measured and theoretical values are in good agreement at a SLD of 4 m. Overall, STI values in Fig. 3 range between 0.63 and 0.8 as the SNR ranges from +8 dB to +∞ (i.e., quiet), a range denoting "good" to "excellent" signal quality for listeners with normal hearing (Houtgast *et al.*, 1980). An example of this signal quality is illustrated by Payton and Braida (1999), who showed that the 0.63–0.8 STI range corresponded to an (approximate) intelligibility range of 88% correct to 96% correct[1] for key words in unprocessed nonsense sentences in reverberation and/or noise. Similarly, the present study's average intelligibility scores for unprocessed speech remained above 90% correct for SNR≥ +8 dB. For vocoded speech, the same increases in SNR and STI are associated with substantial increases in intelligibility (on average, 24%), with the rate of increase rising sharply as the number of channels increases. The largest increase (54%) is observed for the 12-channel vocoder, a configuration for which neither floor nor ceiling effects were observed.

In contrast, SLD has much less influence on STI and intelligibility than SNR. Increases in SLD from 1 to 4 m were associated with decreases of only 0.04 in average STI and 7.4% in the average intelligibility score. The relatively minor effects of SLD on STI are expected, and may be attributed in part to the small size of the room and to placement of the three listening positions in the room's reverber-

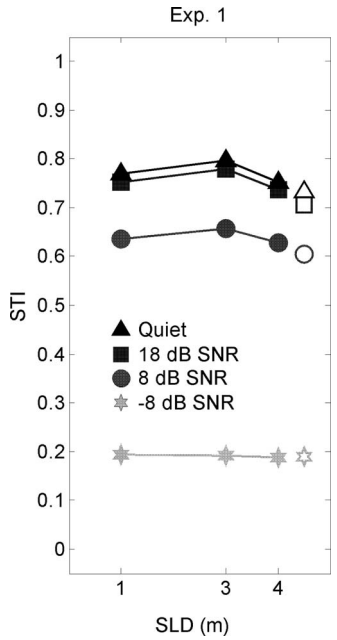


FIG. 3. STIs for the listening conditions of experiment 1 in quiet and in speech-spectrum noise as a function of SLD. Unfilled symbols depict predictions of STI values for reverberant-field listener locations as modeled by Eq. (5).

TABLE II. *A*-weighted SPLs for direct, early, and late arrivals in experiment 1 listening conditions.

SLD (m)	Direct	Early	Late
1	60.5	62.4	56.1
3	52.3	64.2	55.6
4	51.3	63.6	57.9

ant field. A small increase of 0.026 in average STI is also observed as SLD increases from 1 to 3 m: This increase is associated with increased intelligibility scores at 3 m in quiet and at +18 dB SNR.

The coincidence of higher intelligibility scores and higher STI values at the 3 m SLD warranted further investigation. Toward this end, the impulse response filters for each listener position were partitioned into three smaller filters: one producing direct sound arrivals (time span: 0–3 ms), one producing early reflections (time span: 3–50 ms), and one producing later reflections (time span: 50–500 ms). The filters were each convolved with a 15-s recording of speech and *A*-weighted rms output levels were recorded for each filter's output. The *A*-weighted levels of the direct, early, and late portions of the received signal are shown below in Table II. At the 1 m SLD, direct and early arrivals are nearly equal in level, and are each at least 4 dB higher than late arrivals, which may be likened to additive noise in this situation (Lochner and Burger, 1964). At the 3 m SLD, early reflections increase in level while late reflections decrease slightly; providing an 8.6 dB early-to-late reflection ratio. The rms direct arrival level is approximately 3 dB below the late reflection "noise floor," with some direct arrivals likely audible. The combination of strong early reflections and potentially audible direct arrivals is associated with a small increase in performance for implant simulations. At the 4 m SLD, the early-to-late ratio decreases to 5.7 dB, while the direct level drops more than 6 dB below the late reflection level. The combination of these two decreases is associated with a decrease in performance for implant simulations. In contrast, performance for unprocessed speech is not affected by SLD or the changes in early-to-late reflection ratio that are associated with SLD. These findings suggest that simulated implant users may have some limited ability to utilize early reflections, albeit only in the presence of detectable direct arrivals. This dependence on direct arrivals will be explored further in experiment 2.

3. Relationship between STI values and intelligibility scores

The similarities between trends in STI and intelligibility data suggest a strong association between STI and intelligibility scores. This association is depicted in Fig. 4, which plots the intelligibility scores for each channel configuration as a function of STI. The data obtained by Poissant *et al.* (2006) for these listening conditions are plotted for reference. To better explore this association, percent-correct intelligibility curves for each channel configuration were initially fit by sigmoid functions of the form

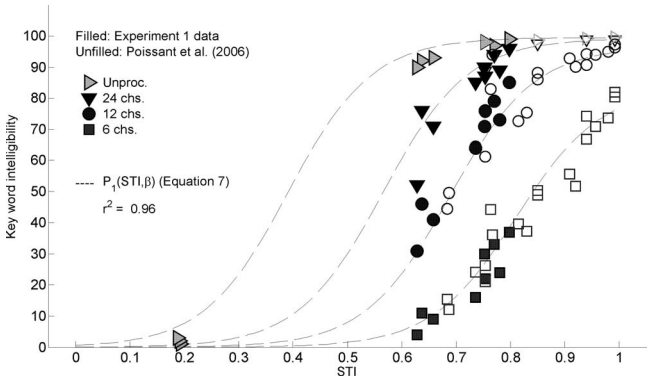


FIG. 4. Speech recognition performance for 6-, 12-, and 24-channel vocoded speech and unprocessed/natural speech in the simulated room of experiment 1 and Poissant *et al.* (2006) as a function of STI. Data from experiment 1 are represented by filled symbols: data from Poissant *et al.* (2006) are represented by unfilled symbols. The dashed line depicts predicted recognition performance as modeled by Eq. (7).

$$P_0(\mathrm{STI}) = \frac{100}{1 + 10^{-[C_1(\mathrm{STI}-C_2)+C_3]}}, \tag{6}$$

where C_1, C_2, and C_3 were fitting constants corresponding to the slope, x-value, and y-value of a reference point on the curve. (C_3=0 for reference points at the 50%-correct level.) These initial attempts revealed logarithmic relationships between both C_1 and C_2 and the channel bandwidth as measured in units of "Cams" (i.e., equivalent-rectangular bandwidth; Glasberg and Moore, 1990). This observation led to the use of a sigmoid function model with logarithmically varying fitting coefficients

$$P_1(\mathrm{STI},\beta) = \frac{P_{\max}()}{1 + 10^{-[(5.59-0.04\ln\beta)((\mathrm{STI}-0.43-0.18\ln\beta)-0.58)]}}, \tag{7}$$

where β represented the channel bandwidth in Cams (i.e., 27.92 Cams divided by the number of channels), $P_{\max}(\beta)$ was the maximum intelligibility score measured for a vocoder with channel bandwidth of β, and the number of available channels for unprocessed speech was assumed to equal 64 (Shannon *et al.*, 2004). $P_{\max}(\beta)$ data from both studies were described well by the equation

$$P_{\max}(\beta) = 99.2 + 1.179\beta - 0.977\beta^2. \tag{8}$$

The good agreement between the data and Eq. (7) suggests that preserving the fidelity of envelope modulations (and thus maximizing STI) will result in the best possible intelligibility.

III. EXPERIMENT 2: EFFECTS OF DISTANCE AND ROOM ABSORPTION ON INTELLIGIBILITY OF PROCESSED SENTENCES

The results of experiment 1 indicate that simulated implant users seated near the front of a small classroom will have speech recognition scores that are slightly (but not significantly) higher than scores for users in the reverberant field of the classroom. The differences between positions are smallest for listeners in quiet with 12 or more spectral channels, and substantially larger for listeners with fewer chan-

nels available and/or noise present. Listeners with normal hearing exhibited no difficulties in any position, presumably because they were able to make better use of early reflections than the simulated implant users. This finding suggests that simulated implant performance will be greatest only when both direct arrivals and early reflections are much higher in level than the reverberant field. The purpose of experiment 2 was to test this hypothesis with simulated implant users in a larger classroom containing listener positions with both positive and negative direct-to-reverberant energy ratios.

A. Methods

1. Subjects

Nine adult listeners (eight females and one male) participated in experiment 2. The subjects' ages ranged from 22 to 38 years (mean age=26.4 years). All of the subjects were native speakers of American English with normal hearing (thresholds $\leq$20 dB HL). None of the subjects had participated in previous CI simulation experiments. All subjects were paid for their participation.

2. Materials and processing

Processing for experiment 2 was similar to that of experiment 1, with noteworthy changes in noise, reverberation, and vocoder conditions.

Noise processing. The 360-sentence recordings of experiment 1 were input to the simulators either as recorded in quiet or with either the speech-spectrum noise of experiment 1 or two-talker babble added at a SNR of 18 dB. The two-talker babble was the same as that used in Poissant *et al.*, 2006, derived from digital recordings of two college-aged female students speaking different sets of syntactically correct nonsense sentences. Pauses between sentences were removed to produce two recordings of continuous speech, which were then matched in rms level and combined to produce two-talker babble.

Reverberation simulation. The reverberation simulator of experiment 1 was used to model an idealized rectangular classroom ($6.7\times10.1\times2.6$ m^3) with each of the four values of α (1.0, 0.7, 0.4, and 0.25) used in Poissant *et al.*, 2006. The dimensions for the ideal classroom were taken from a real rectangular classroom at the University of Massachusetts Amherst. The details of the room simulation are identical to those of Fig. 1 with several notable exceptions: the larger room width and length, the distance between the listener position and most distant wall (increased proportionally to 2.8 m), and selected SLDs of 1, 4, or 7 m. The 1000 Hz octave-band RT_{60} values for the three positions (measured as in experiment 1) were approximately 680 ms for α=0.25, 380 ms for α=0.4, and 170 ms for α=0.7. Theoretical critical distances and measured direct-to-reverberant energy ratios for each position in the three reverberant rooms are shown in Table III. The chosen combinations of SLDs and α values place the 1 m SLD position near the theoretical critical distance for α=0.25, and within the critical distance for α=0.4 and 0.7. The 4 and 7 m positions remain in the reverberant field for $\alpha<1$.

TABLE III. Direct-to-reverberant energy ratios (in decibels) for experiment 2 listening conditions in reverberant rooms.

SLD (m)	α 0.25	0.40	0.70
1	−0.62	2.31	7.55
4	−11.93	−8.57	−2.04
7	−15.49	−11.95	−4.68
d_c (m)	1.05	1.33	1.75

As in experiment 1, all speech signals were presented to the subjects at 65 dBA. Unnormalized speech levels at each position for a 65 dBA direct field at 1 m are shown in Table IV. Level differences were, as in experiment 1, consistent with theoretical predictions (Barron and Lee, 1988; Sato and Bradley, 2008). It should be noted that without SPL normalization listeners in rooms with α=0.7 or α=1.0 could receive speech at levels near 50 dB SPL, a level that has been shown to impair intelligibility for implant users (Skinner *et al.*, 1997; Firszt *et al.*, 2007). Since large interstimulus presentation level differences could confound effects of early reflections, we chose to present all signals at the same level as in experiment 1. Actual implant users would have similar capabilities to compensate for these differences by manually adjusting microphone sensitivity (Donaldson and Allen, 2003; James *et al.*, 2003) and/or using adaptive dynamic range optimization or other automatic gain control algorithms to amplify soft speech (James *et al.*, 2002; Dawson *et al.*, 2004).

Implant simulation. The reverberated sentences were processed by only the six-channel vocoder of experiment 1, which, in previous work (Whitmal *et al.*, 2007), was determined to correspond to the effective number of channels many CI listeners can access (Dorman and Loizou, 1998), and represents a configuration for which simulation results are similar to results from CI systems (Dorman *et al.*, 1998; Friesen *et al.*, 2001).

3. Procedure

Testing procedures for experiment 2 were similar to those of experiment 1, with one notable change: combinations of the three noise conditions (quiet, speech-spectrum noise at +18 dB SNR, and two-talker babble at +18 dB SNR), four absorption coefficients, and three SLDs were used to process each of the 36-sentence lists. Practice materials consisted of 40 sentences, divided into 8 groups of 5 and presented without feedback at the beginning of the experiment. Each group of practice sentences was processed by a unique combination of two SLDs (1 and 7 m), two absorp-

TABLE IV. *A*-weighted SPLs for experiment 2 listening conditions when the direct field SPL at 1 m equals 65 dBA.

SLD (m)	α 0.25	0.40	0.70	1.00
1	68.5	67.2	66.0	65.0
4	64.9	62.0	57.2	53.0
7	63.7	60.0	53.7	48.1

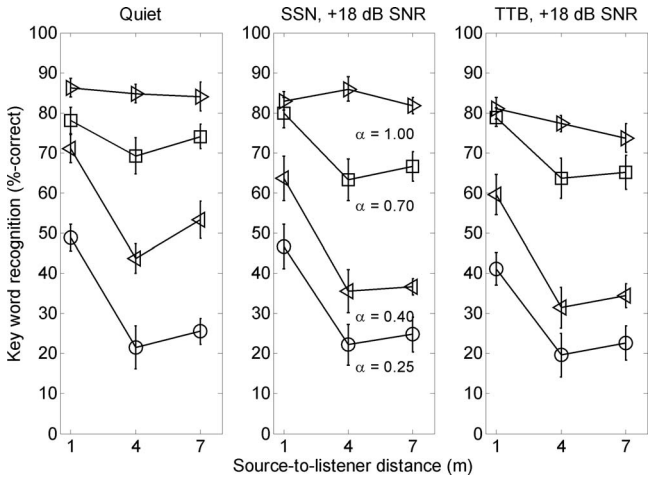


FIG. 5. Speech recognition performance for six-channel vocoded speech in quiet (left panel), speech-spectrum noise (center panel), and two-talker babble (right panel) as a function of SLD and room absorption coefficient in the experiment 2 room simulation. Error bars represent ±1 standard error.

tion coefficients (α=1.0 and 0.4), and two noise conditions (quiet and speech-spectrum noise at +18 dB SNR). The practice sentences were not used in the main experiment.

B. Computation of STI values

STI computations for experiment 2 followed the procedures used for experiment 1.

C. Results

1. Intelligibility scores

Intelligibility scores for experiment 2 were derived from the percentage of correctly repeated key words per condition. Mean intelligibility scores for listening in quiet and in the two noise conditions are shown in Fig. 5. Scores in quiet were strongly dependent on α, with the average score at 1 m decreasing from 82.3% correct to 48.9% correct as α decreased from 1.00 to 0.25. Intelligibility scores in quiet also decreased as the SLD increased from 1 to 4 m, with average decreases of 8.9%, 27.4%, and 27.4% observed for α=0.7, 0.4, and 0.25, respectively. In general, average scores at 7 m were approximately equal to average scores at 4 m. For α=0.4 in quiet, average scores at 7 m (53.33% correct) were higher than scores at 4 m (43.70% correct), an advantage that, while unexpected, is not statistically significant.

Average scores in noise were lower than average scores in quiet, with the degree of difference determined by values of α. For α=1.0, average scores for the speech-spectrum noise condition were approximately equal to scores in quiet, and 6.33% higher than scores for the two-talker babble condition. For $\alpha<1.0$, average speech-spectrum noise scores were approximately equal to average two-talker babble scores. The advantage observed for speech-spectrum noise in anechoic conditions and the equivalence of speech-spectrum noise and two-talker babble in reverberant conditions are both consistent with results from previous work (Poissant *et al.*, 2006; Whitmal *et al.*, 2007). As α decreased, the differences between scores in quiet and corresponding scores in noise also decreased, such that scores in quiet and in noise were approximately equal when α=0.25.

Subject scores were converted to rationalized arcsine units (Studebaker, 1985) and input to a repeated-measures analysis of variance of intelligibility scores. All within-subject factors for the analysis of variance were statistically significant; these included α ($F[3,236]=460.24$, $p<0.0001$), SLD ($F[2,236]=89.49$, $p<0.0001$), and noise condition ($F[2,236]=16.65$, $p<0.0001$). The first-order interaction between α and SLD ($F[6,236]=11.98$, $p<0.0001$) was significant, reflecting the tendency of intelligibility in the reverberant field to worsen as α increased. The first-order interaction between α and noise condition ($F[6,236]=2.24$, $p=0.04$) was also significant, reflecting the tendency for speech-spectrum noise scores to match scores in quiet for α=1.0 and match two-talker babble scores for $\alpha<1.0$. *Post hoc* tests using the Tukey honestly significant difference criterion at the 0.05 level indicated that (a) scores at 1 m were significantly higher than scores at 4 or 7 m, (b) scores at 4 and 7 m were not significantly different from each other, (c) scores for each α value were all significantly different from each other, and (d) scores for each noise condition were all significantly different from each other.

To further explore the significance of adding reverberation, four *post hoc* analyses of variance of scores obtained for individual values of α were conducted. These analyses indicated that (a) for α=1.0, scores for quiet and speech-spectrum noise were significantly greater than scores for two-talker babble; (b) for $\alpha<1.0$, there were no significant differences between speech-spectrum noise and two-talker babble; (c) for α=0.25 and 0.7, scores in quiet were not significantly different from scores in speech-spectrum noise or two-talker babble; and (d) for α=1.0, SLD was not a significant factor.

2. STI values

STI values for the listening conditions of experiment 2 are shown in Fig. 6. As with experiment 1, measured and theoretical values are represented by filled and unfilled symbols, respectively. Two other similarities between these data and those of experiment 1 are evident. First, the measured and theoretical STI values of Fig. 6 are also in good agreement for positions in the reverberant field (SLD$\geq$4 m). Second, the range of experiment 2 STI values (0.63–0.99) also reflects good signal quality, and increases in STI over this range (whether produced by changes in SNR or α) are likewise associated with increases in intelligibility. Unlike experiment 1, however, the larger room volume and variable range of α values used in experiment 2 enabled changes in SLD to have a greater effect on STI (and envelope modulations) than changes in SNR. When SLD was increased from 1 to 4 m, the average STI decreased by 0.068 for α=0.4 or 0.7, with intelligibility decreases of 27.4% and 8.9%; for α=0.25, average STI decreased by 0.055 and intelligibility decreased by 27.4%. These changes in STI (and intelligibility) are consistent with trends observed in an idealized model of direct field contributions to the STI (Houtgast *et al.*,

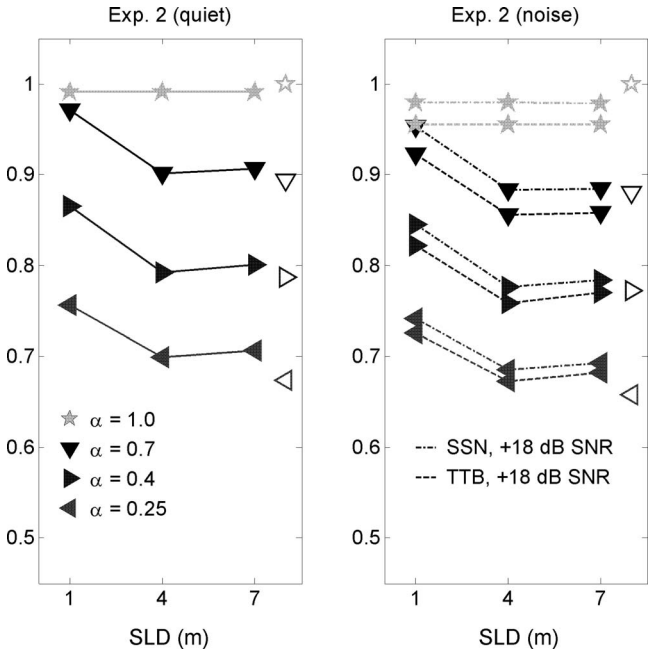


FIG. 6. (a) STIs for the listening conditions of experiment 2 in quiet as a function of SLD and room absorption. (b) STIs for the listening conditions of experiment 2 in speech-spectrum noise (dashed/dotted line) and in two-talker babble (dashed line) as a function of SLD and room absorption. Unfilled symbols depict predictions of STI values for reverberant-field listener locations as modeled by Eq. (5).

1980). Adding speech-spectrum noise to the speech decreased average STI values by less than 0.020 and intelligibility scores by 4.2%; adding two-talker babble to the speech decreased average STI values by between 0.036 and 0.048 and intelligibility scores by 7.4%. The larger STI decrease associated with adding two-talker babble is presumably caused by the envelope modulations of the babble, which act to reduce the average modulation depth attributable to the target speech.

The impulse response filters for each listener position were partitioned into three smaller filters as in experiment 1. *A*-weighted levels of the direct, early, and late portions of the received signal for each combination of SLD and α are shown in Table V. For α=0.25, direct field intensity drops below both early and late reflection intensities as SLD increases; intelligibility is best at the 1 m SLD, where the direct field is stronger than both the early and late reflections. Similar patterns are apparent for α=0.40 and α=0.70, albeit with both higher intelligibility and higher direct-to-reverberant energy ratios observed at all SLDs.

3. Relationship between STI values and intelligibility scores

The relationship between STI and intelligibility data for experiment 2 is depicted in Fig. 7, along with a best-fit modified sigmoid curve (dashed line)

$$P_2(\mathrm{STI},\beta) = \frac{P_{\max}(\beta)}{1 + 10^{-[(3.08+1.82\ln\beta)((\mathrm{STI}+0.005-0.50\ln\beta)+0.02)]}}. \quad (9)$$

The good agreement between the data and Eq. (9) suggests that the effect of increasing SLD is, like noise and reverberation, manifested as distortions in envelope modulation that reduce intelligibility. Figure 7 also displays the best-fit line for experiment 1 data (dotted-dashed line), which falls (on average) 7.9% below the line for experiment 2. This large difference may be attributable in part to differences in experimental conditions and subject acclimatization. Although the subjects for each experiment listened to the same sentences, experiment 1 subjects were confronted with more adverse conditions (SNR=−8 and +8 dB) than experiment 2 subjects. Moreover, experiment 1 subjects listened to four different vocoders while experiment 2 subjects listened to only one. Reducing the time that experiment 1 subjects listened to the six-channel vocoder under favorable conditions may have prevented them from acclimating to the vocoder as well as the experiment 2 subjects did.

TABLE V. *A*-weighted SPLs for direct, early, and late arrivals in experiment 2 reverberant listening conditions.

SLD (m)	Direct	Early	Late
	α=0.25		
1	61.4	60.8	56.2
4	52.7	63.2	58.3
7	49.6	63.4	58.9
	α=0.4		
1	62.5	60.0	50.8
4	55.4	63.5	55.9
7	53.0	64.0	55.9
	α=0.7		
1	63.7	56.4	36.8
4	60.3	62.1	44.8
7	59.7	63.6	44.8

IV. DISCUSSION

A. Effects of SLD on intelligibility

For speech processed in a way that makes it vulnerable to the effects of reverberation (i.e., when it is vocoded with a restricted number of spectral channels), we found that SLD can matter a great deal to speech understanding in a large room. The degree to which benefits of reductions in SLD will be realized will depend on the size of the room and the levels of reverberation and ambient noise, as well as whether or not the separation between source and listener is within (or close to) the critical distance of the room. In experiment 1, subjects demonstrated a modest benefit from reducing SLD in each vocoder configuration in at least some conditions (e.g., those that did not produce ceiling or floor effects). This finding is likely a result of the fact that the signal reaching them was comprised largely of early reflections, rather than direct sound. Experiment 2, conducted within a larger simulated room, produced results that demonstrated a much more promising effect of decreasing SLD on understanding of CI processed speech in reverberant spaces. Again, changes in distance that kept the listener rather deep in the reverberant field (i.e., from 7 to 4 m) had no positive impact on performance. However, once the listener's position crossed

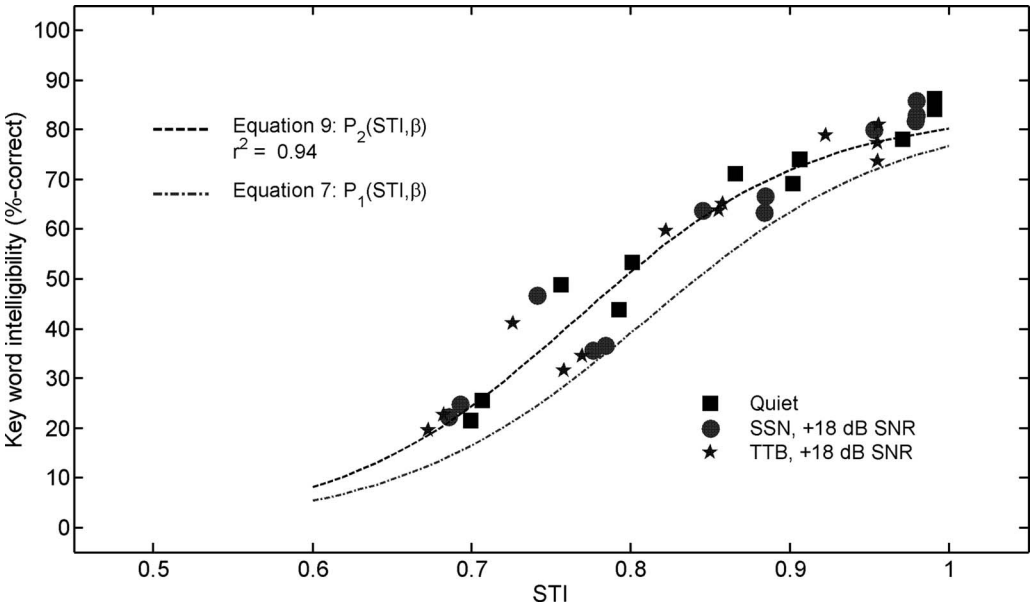


FIG. 7. Speech recognition performance for six-channel vocoded speech in the simulated room of experiment 2. The dashed line depicts predicted recognition performance as modeled by Eq. (9); the dotted-dashed line depicts predicted recognition performance as modeled by Eq. (7) for experiment 1 data.

from the reverberant to the direct field (or even close to the direct field), very large improvements were realized. In quiet, the extent of SLD-produced improvements was commensurate with the differences in direct-to-reverberant energy ratio for the 1 and 4 m SLDs. This is evident in Fig. 5, which indicates large improvements in intelligibility for $\alpha \leq 0.40$ when SLD was reduced from 4 to 1 m, a smaller improvement for $\alpha=0.70$, and no significant (or expected) improvement for $\alpha=1.0$. In all cases, the improved scores were significantly lower than scores observed at 1 m for $\alpha=1.0$ (see Fig. 7), indicating that the effects of reverberation could not be completely eliminated. These findings underscore our inability to fully compensate for the detrimental effects of reverberation simply by reducing the SLD.

B. STI-based predictions of intelligibility

The STI has been shown to be an accurate and reliable predictor of speech intelligibility in reverberation and noise for normal-hearing listeners. For this reason, the STI and variants such as the rapid STI (Steeneken and Houtgast, 1982) or the STIPA metric for public address systems (Steeneken *et al.*, 2001; Bjør, 2004) have been widely used in the assessment of classrooms and sound reinforcement systems. In particular, several investigators (Bradley, 1986b; Bradley *et al.*, 1999; Siebein *et al.*, 2000; Crandell *et al.*, 2004) have advocated using STI values (computed from actual room impulse response measurements) to evaluate the suitability of classrooms and identify problems for remediation. Other investigators have used the STI to guide the design of simulated classrooms that optimize speech intelligibility (Bistafa and Bradley, 2000, 2001; Yang and Hodgson, 2006).

The present results suggest that the STI can also be used to accurately predict intelligibility in simulations of CI processing. The uses of the STI in room design and assessment described above may therefore also be applicable to CI simulations. Moreover, the good agreement observed between predicted STI values [computed via Eq. (5)] and measured STI values further suggests that the quality of seating in the reverberant field in a room can be estimated simply for simulated CI users if the volume, reverberation time, and SNR are known. Seating in the direct field poses a greater challenge for such simple estimates, since closed-form equations modeling room acoustics for the STI (e.g., Houtgast *et al.*, 1980) are unable to accurately model effects of early reflections.

C. Comparisons with previous vocoder experiments

One goal of the present study was to extend the results of Poissant *et al.* (2006) concerning reverberation and noise effects on vocoded speech intelligibility. This previous study included two experiments that used the room model of the present experiment 1 with a SLD of 4 m and various values of α. Mean scores in quiet for the present experiment 1 and for the first experiment of Poissant *et al.* (2006) at 4 m with $\alpha=0.25$ were nearly identical. The second experiment of Poissant *et al.* (2006) examined performance in quiet and noise as a function of α for 6-channel and 12-channel vocoders. Their data (shown in the left panel of Fig. 8) indicated that (a) intelligibility increased in a near-linear fashion as α increased from 0.4 to 1.0, (b) the effects of speech-spectrum noise and two-talker babble on intelligibility were not significantly different, and (c) scores for all conditions increased at the approximate rate of 5% per 0.1 increase in α without any significant interaction between the number of channels, noise level, and α. As a result, the effects of reverberation and noise on intelligibility appeared to be additive, a result that differed markedly from those of previous studies showing interactions between noise and reverberation for unprocessed speech (Nabelek and Mason, 1981; Loven and Collins, 1988; Helfer and Wilber, 1990; Payton *et al.*, 1994).

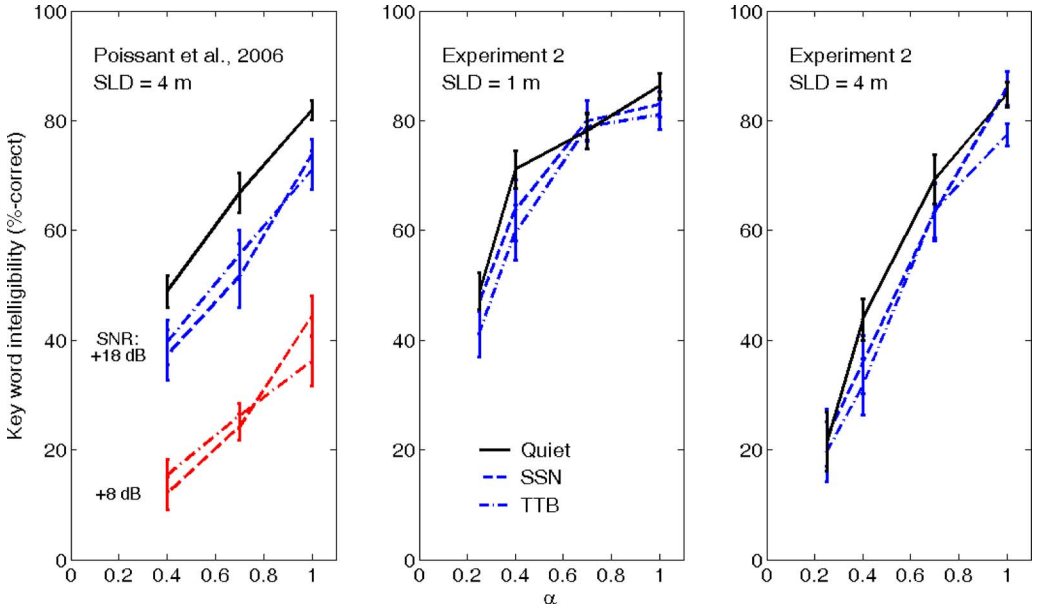


FIG. 8. (Color online) Speech recognition performance for six-channel vocoded speech in experiment 2 of Poissant *et al.* (2006), experiment 2 of the present work at a 1 m SLD (center panel), and a 7 m SLD (right panel) as a function of room absorption coefficient. Error bars represent ±1 standard error.

In the present experiment 2, performance was evaluated in a larger room for a six-channel vocoder at three different SLDs. Intelligibility scores for the six-channel vocoder are plotted in Fig. 8 as functions of α and SNR for SLD=1 m (center panel) and SLD=4 m (right panel); the patterns of results for 7 m mirrored that of the 4 m data and therefore the 7 m data are not presented. Scores in quiet are similar in value to those measured in experiment 2 of Poissant *et al.* (2006), as are scores in noise at α=0.4 and 1.0. Moreover, the curves for scores in quiet and at +18 dB SNR are nearly parallel, which, aside from ceiling and floor effects, reflects the same limited interaction observed by Poissant *et al.* (2006).

D. Implications for listeners with CIs

The present study used channel vocoders to model speech perception of CI users in reverberant environments. Numerous studies (e.g., Dorman and Loizou, 1998; Friesen *et al.*, 2001) have shown that experiments with channel vocoders can approximate best-case performance for implant users. To the extent that vocoded speech resembles implant processed speech, the present results suggest that intelligibility for implant users can improve when SLD is reduced. While modest improvement may occur as SLD decreases within the reverberant field, striking and important improvements are seen when distance decreases enough to place the listener within, or even close to, the direct field. These effects seem to be well captured by the STI (as shown in a comparison between Figs. 5 and 7), which may prove to be a helpful predictive tool. In everyday listening environments, this means that room size will be a key factor in determining potential benefit from preferential seating. For example, in the smaller classroom modeled in experiment 1, the listener remained in the reverberant field even when seated 1 m from the sound source. It is unlikely that CI users will easily be able to position themselves closer than 1 m to most talkers. In the larger modeled room of experiment 2, listeners seated at a 1 m SLD were within (or very close to) the direct field, whereas listeners seated at a 2 m SLD would have been within the reverberant field. As a result, merely reducing the SLD from 2 to 1 m would be expected to be associated with large improvements in performance.

It is also important to note that the intelligibility gained by reducing the SLD in rooms with low-to-moderate levels of absorption did not fully compensate for the intelligibility lost to reverberation (see Fig. 7). These findings highlight the very important negative impact of reverberation and suggest that one of the first steps in improving speech understanding for a CI user in a real-world environment (e.g., a classroom or a conference room) should be to reduce the level of reverberation to the greatest extent possible. However, as there are limits to our ability to ensure appropriate levels of reverberation in all environments in which a CI user may converse, we must remain ever cognizant of making recommendations for preferential seating as we have demonstrated that reliance on decreasing SLD will prove beneficial in improving speech understanding, particularly if the listener is able to move within the critical distance of the room. Another very important and perhaps fail-safe strategy for improving intelligibility of CI users in real-world environments is the use of an FM system or similar remote assistive listening device to replace the reverberated signal reaching the listener with a clean near-field signal. It is also important to remember that only the smearing aspect of reverberation was simulated in the present investigation; the rms levels across conditions were equated. This was done in order to allow for an investigation of the specific impact of temporal smearing on intelligibility in listeners who essentially have access to only envelope cues. At the same time, however, it means that the present data are not influenced by any increase in level that would typically accompany reverberation as well as any in-

creases in intensity that occurs when SLD decreases. Such increases could potentially improve performance for some CI users.

The present study considered only speech and noise arriving coincidentally from the same source. In typical settings it is likely that speech and noise would be spatially distinct, allowing individuals with normal hearing to make use of binaural processes to separate the target from the masker. While a changing trend toward bilateral implantation is observable, the majority of CI users have received just one implant. As unilateral listeners, their best hope of benefit from spatially separated signal and noise sources is a head shadow effect when the noise is located on the side of their head opposite from their implant. Further, for those users who are bilaterally implanted, uncoordinated input to the two clinical speech processors could impose limitations on the use of cues that would result in the ability to suppress noise in favor of speech (i.e., the "binaural squelch" effect) and other expected speech-in-noise improvements afforded to listeners with binaural hearing (MacKeith and Coles, 1971). Finally, it is also possible that CI users might benefit from spatial differences in early reflection patterns, which have been shown to increase the effective signal level and improve intelligibility in some situations (Barron and Lee, 1988; Yang and Hodgson, 2006; Sato and Bradley, 2008). This possibility will be explored in future investigations.

V. SUMMARY AND CONCLUSIONS

Speech intelligibility for listeners using CIs can be compromised by the temporal smearing effects of reverberation and additive noise. For vocoder simulations in normally hearing listeners, the effects of both factors can be reduced substantially by moving the listener as close as possible to the speech source provided that the reduction in distance places the listener in or very near the direct field. The present study investigated the potential benefits of this strategy for CI users in two simulated classrooms. Results of the study suggest that CI users should receive some benefit by moving closer to the speech source in large rooms with low-to-moderate absorption. Limited benefits are expected in small rooms where all listener positions are in the reverberant field.

ACKNOWLEDGMENTS

We would like to thank Kristina Curro for help in testing subjects, and Dr. Richard Freyman for his support and helpful suggestions. We would also like to thank Associate Editor Ruth Litovsky and three anonymous reviewers for their helpful suggestions. Funding for this research was provided by the National Institutes of Health (NIDCD Grant No. R03 DC7969).

[1]Estimated from inspection of Fig. 10 (bottom panel) of Payton and Braida (1999). The reverberation used by those authors was produced by the same image-source algorithm as used in the present work.

Allen, J. B., and Berkley, D. A. (**1979**). "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943–950.

Anderson, K. A., Goldstein, H., Colodzin, L., and Iglehart, F. (**2005**). "Benefit of S/N enhancing devices to speech perception of children listening in a typical classroom with hearing aids or a cochlear implant," J. Educ. Audiol. **12**, 14–28.

ANSI (**1969**). *ANSI S3.5–1969: American National Standards Methods for the Calculation of the Articulation Index* (American National Standards Institute, New York).

ANSI (**1997**). *ANSI S3.5–1997: American National Standards Methods for the Calculation of the Speech Intelligibility Index* (American National Standards Institute, New York).

ANSI (**2002**). *ANSI S12.60-2002: Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools* (American National Standards Institute, New York).

Barron, M., and Lee, L.-J. (**1988**). "Energy relations in concert auditoriums. I.," J. Acoust. Soc. Am. **84**, 618–628.

Bistafa, S. R., and Bradley, J. S. (**2000**). "Reverberation time and maximum background-noise level for classrooms from a comparative study of speech intelligibility metrics," J. Acoust. Soc. Am. **107**, 861–875.

Bistafa, S. R., and Bradley, J. S. (**2001**). "Predicting speech metrics in a simulated classroom with varied sound absorption," J. Acoust. Soc. Am. **109**, 1474–1482.

Bjør, O.-H. (**2004**). "Measure speech intelligibility with a sound level meter," Sound Vib. **38**, 10–13.

Bolt, R. H., and MacDonald, A. D. (**1949**). "Theory of speech masking by reverberation," J. Acoust. Soc. Am. **21**, 577–580.

Bradley, J. S. (**1986a**). "Predictors of speech intelligibility in rooms," J. Acoust. Soc. Am. **80**, 837–845.

Bradley, J. S. (**1986b**). "Speech intelligibility studies in classrooms," J. Acoust. Soc. Am. **80**, 846–854.

Bradley, J. S., Reich, R. D., and Norcross, S. G. (**1999**). "On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility," J. Acoust. Soc. Am. **106**, 1820–1828.

Bradley, J. S., Sato, H., and Picard, M. (**2003**). "On the importance of early reflections for speech in rooms," J. Acoust. Soc. Am. **113**, 3233–3244.

Crandell, C., Holmes, A., Flexer, C., and Payne, M. (**1998**). "Effects of sound field FM amplification on the speech recognition of listeners with cochlear implants," J. Educ. Audiol. **6**, 21–27.

Crandell, C., Kreisman, B. M., Smaldino, J., and Kreisman, N. V. (**2004**). "Room acoustics intervention efficacy measures," Semin. Hear. **25**, 201–206.

Crandell, C., and Smaldino, J. (**2000**). "Classroom acoustics for children with normal hearing and with hearing impairment," Lang. Spch. Hear. Svcs. In Schools **31**, 362–370.

Culling, J. F., Hodder, K. I., and Toh, C. Y. (**2003**). "Effects of reverberation on perceptual segregation of competing voices," J. Acoust. Soc. Am. **114**, 2871–2876.

Dawson, P. W., Decker, J. A., and Psarros, C. E. (**2004**). "Optimizing dynamic range in children using the nucleus cochlear implant," Ear Hear. **25**, 230–241.

Donaldson, G. S., and Allen, S. L. (**2003**). "Effects of presentation level of phoneme and sentence recognition in quiet by cochlear implant listeners," Ear Hear. **24**, 392–405.

Dorman, M. F., and Loizou, P. C. (**1998**). "The identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations of processors with two to nine channels," Ear Hear. **19**, 162–166.

Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z. (**1998**). "The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels," J. Acoust. Soc. Am. **104**, 3583–3585.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," J. Acoust. Soc. Am. **102**, 2403–2411.

Dreschler, W. A., and Leeuw, A. R. (**1990**). "Speech reception in reverberation related to temporal resolution," J. Speech Hear. Res. **33**, 181–187.

Duquesnoy, A. J., and Plomp, R. (**1980**). "Effect of reverberation and noise on the intelligibility of sentences in cases of presbyacusis," J. Acoust. Soc. Am. **68**, 537–544.

Finitzo-Hieber, T., and Tillman, T. W. (**1978**). "Room acoustics effects on monosyllabic word discrimination ability for normal and hearing-impaired children," J. Speech Hear. Res. **21**, 441–458.

Firszt, J. B., Holden, L. K., Skinner, M. W., Tobey, E. A., Peterson, A., Gaggi, W., Runge-Samuelson, C. L., and Wackym, P. A. (**2004**). "Recognition of speech presented at soft to loud levels by adult cochlear implant

recipients of three cochlear implant systems," Ear Hear. **25**, 375–387.
Francis, W. N., and Kucera, H. (**1982**). *Frequency Analysis of English Usage: Lexicon and Grammar* (Houghton Mifflin, Boston, MA).
Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.
Fu, Q.-J., and Shannon, R. V. (**2000**). "Effect of stimulation rate on phoneme recognition by nucleus-22 cochlear implant listeners," J. Acoust. Soc. Am. **107**, 589–597.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**,103–138.
Goldsworthy, R. L., and Greenberg, J. E. (**2004**). "Analysis of speech-based speech transmission index methods with implications for nonlinear operations," J. Acoust. Soc. Am. **116**, 3679–3689.
Helfer, K. S. (**1994**). "Binaural cues and consonant perception in reverberation and noise," J. Speech Hear. Res. **37**, 429–438.
Helfer, K. S., and Freyman, R. L. (**2004**). "Development of a topic-related sentence corpus for speech perception research," J. Acoust. Soc. Am. **115**, 2601–2602.
Helfer, K. S., and Freyman, R. L. (**2005**). "The role of visual speech cues in reducing energetic and informational masking," J. Acoust. Soc. Am. **117**, 842–849.
Helfer, K. S., and Wilber, L. (**1990**). "Hearing loss, aging, and speech perception in reverberation and noise," J. Speech Hear. Res. **33**, 149–155.
Houtgast, T., and Steeneken, H. J. M. (**1973**). "The modulation transfer function in acoustics as a predictor of speech intelligibility," Acustica **28**, 66–74.
Houtgast, T., and Steeneken, H. J. M. (**1985**). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**, 1069–1077.
Houtgast, T., Steeneken, H. J. M., and Plomp, R. (**1980**). "Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics," Acustica **46**, 60–72.
IEC (**2002**). *Publication IEC 60268-16: Sound System Equipment—Part 16. Objective Rating of Speech Intelligibility by Speech Transmission Index*, 3rd ed. (International Electrotechnical Commission, Geneva, Switzerland).
Iglehart, F. (**2004**). "Speech perception by students with cochlear implants using sound-field systems in classrooms," Am. J. Audiol. **13**, 62–72.
James, C. J., Blamey, P. J., Martin, L., Swanson, B., Just, Y., and Macfarlane, D. (**2002**). "Adaptive dynamic range optimization for cochlear implants: A preliminary study," Ear Hear. **23**, 49S–58S.
James, C. J., Skinner, M. W., Martin, L. F. A., Holden, L. K., Galvin, K. L., Holden, T. A., and Whitford, L. (**2003**). "An investigation of input level range for the nucleus 24 cochlear implant system: Speech perception performance, program preference, and loudness comfort ratings," Ear Hear. **24**, 157–174.
Kuttruff, H. (**1979**). *Room Acoustics* (Taylor and Francis, London).
Latham, H. G. (**1979**). "The signal-to-noise ratio for speech intelligibility—An auditorium acoustics design index," Appl. Acoust. **12**, 253–320.
Lehmann, E. A., and Johansson, A. M. (**2008**). "Prediction of energy decay in room impulse responses simulated with an image-source model," J. Acoust. Soc. Am. **124**, 269–277.
Lochner, J. P. A., and Burger, J. F. (**1964**). "The influence of reflections on auditorium acoustics," J. Sound Vib. **1**, 426–454.
Loven, F. C., and Collins, M. J. (**1988**). "Reverberation, masking, filtering, and level effects on speech recognition performance," J. Speech Hear. Res. **31**, 681–695.
MacKeith, N. W., and Coles, R. R. A. (**1971**). "Binaural advantages in hearing of speech," J. Laryngol. Otol. **85**, 213–232.
Nabelek, A. K., Letowski, T. R., and Tucker, F. M. (**1989**). "Reverberant overlap- and self-masking in consonant identification," J. Acoust. Soc. Am. **86**, 1259–1265.
Nabelek, A. K., and Mason, D. (**1981**). "Effect of noise and reverberation on binaural and monaural word identification by subjects with various audiograms," J. Speech Hear. Res. **24**, 375–383.
Nabelek, A. K., and Robinson, P. K. (**1982**). "Monaural and binaural speech perception in reverberation for listeners of various ages," J. Acoust. Soc. Am. **71**, 1242–1248.
Payton, K. L., and Braida, L. D. (**1999**). "A method to determine the speech transmission index from speech waveforms," J. Acoust. Soc. Am. **106**, 3637–3648.
Payton, K. L., Uchanski, R. M., and Braida, L. D. (**1994**). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," J. Acoust. Soc. Am. **95**, 1581–1592.
Peterson, P. M. (**1986**). "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," J. Acoust. Soc. Am. **80**, 1527–1529.
Poissant, S. F., Whitmal, N. A., and Freyman, R. L. (**2006**). "Effects of reverberation and masking on speech intelligibility in cochlear implant simulations," J. Acoust. Soc. Am. **119**, 1606–1615.
Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.
Sabine, W. C. (**1922**). *Collected Papers on Acoustics* (Harvard University Press, Cambridge, MA).
Sato, H., and Bradley, J. S. (**2008**). "Evaluation of acoustical conditions for speech communication in working elementary school classrooms," J. Acoust. Soc. Am. **123**, 2064–2073.
Schroeder, M. R. (**1965**). "New method of measuring reverberation time," J. Acoust. Soc. Am. **37**, 409–412.
Shannon, R. V., Fu, Q.-J., and Galvin, J. (**2004**). "The number of spectral channels required for speech recognition depends on the difficulty of the listening situation," Acta Oto-Laryngol., Suppl. **124**, 50–54.
Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Siebein, G. W., Gold, M. A., Siebein, G. W., and Ermann, M. G. (**2000**). "Ten ways to provide a high-quality acoustical environment in schools," Lang. Spch. Hear. Svcs. in Schools **31**, 376–384.
Skinner, M. W., Holden, L. K., Holden, T. A., Demorest, M. E., and Fourakis, M. S. (**1997**). "Speech recognition at simulated soft, conversational, and raised-to-loud vocal efforts by adults with cochlear implants," J. Acoust. Soc. Am. **101**, 3766–3782.
Steeneken, H. J. M., and Houtgast, T. (**1982**). "Evaluation of a physical method for estimating speech intelligibility in auditoria," Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. **VII**, pp. 1452–1454.
Steeneken, H. J. M., Verhave, J., McManus, S., and Jacob, K. (**2001**). "Development of an accurate, handheld, simple-to-use meter for the prediction of speech intelligibility," Proceedings of the International Acoustics, UK, Vol. **23**, 53–59.
Stickney, G. S., Zeng, F.-G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.
Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.
Whitmal, N. A., Poissant, S. F., Freyman, R. L., and Helfer, K. S. (**2007**). "Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience," J. Acoust. Soc. Am. **122**, 2376–2388.
Xu, L., and Zheng, Y. (**2007**). "Spectral and temporal cues for phoneme recognition in noise," J. Acoust. Soc. Am. **122**, 1758–1764.
Yang, W., and Hodgson, M. (**2006**). "Auralization study of optimum reverberation times for speech intelligibility for normal and hearing-impaired listeners in classrooms with diffuse sound fields," J. Acoust. Soc. Am. **120**, 801–807.

A simple single-interval adaptive procedure for estimating thresholds in normal and impaired listeners

Wendy Lecluyse and Ray Meddis
Department of Psychology, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, United Kingdom

(Received 8 March 2009; revised 23 June 2009; accepted 1 September 2009)

This report presents a single-interval adaptive procedure for measuring thresholds in untrained normal and impaired listeners. The accuracy of the procedure is evaluated using Monte Carlo methods and human data allowing a method to be proposed for deciding in advance the number of trials required to achieve a specified level of accuracy. The number of trials depends on the slope of the psychometric function. The slope of the psychometric function is evaluated in normal and impaired listeners, and is found to give a useful guide to the required number of trials. The single-interval up/down procedure is subsequently compared with two other popular traditional methods (two-interval forced-choice, two-down/one-up and maximum-likelihood procedures) and is shown to yield similar thresholds and be more efficient.
 [DOI: 10.1121/1.3238248]

PACS number(s): 43.66.Yw [MW] Pages: 2570–2579

I. INTRODUCTION

Researchers who need to make many threshold measurements in both normal and impaired participant groups are confronted with the problem of finding a fast, participant-friendly, and reliable measurement procedure. They must choose between standard clinical methods based on single-interval "yes/no" procedures and those used in psychoacoustic research laboratories based on a multiple-interval forced-choice approach. Clinical methods, such as the modified Hughson–Westlake procedure (Carhart and Jerger, 1959), have been optimized for speed and patient acceptability while laboratory methods aim for greater accuracy and theoretical rigor. The former are simple to administer, require little patient training, can be easily automated, and involve only a small number of trials.

Unfortunately, standard clinical procedures are not acceptable to most of the laboratory-based scientific community because they are believed to overestimate thresholds and fail to accommodate differences in response bias (Marshall and Jesteadt, 1986). This is a problem for the clinical researcher who needs to obtain thresholds that are meaningful within a wider research context where the choice of method is overwhelmingly a multiple-interval, forced-choice approach. Standard laboratory procedures, on the other hand, are more complicated, often require considerable training, and typically need many trials. Opting for the apparently more rigorous laboratory procedure comes at a very high price. The large number of trials (often around 50–60) is a major disincentive. Also, when using these procedures, the patient must choose one of two temporal windows where one is occupied by a stimulus and the other is empty. Patients who may be elderly or have a lower educational level may experience considerable difficulty with this method. The problem is particularly pressing when the stimulus is below threshold and neither window is a straightforward choice. Here participants are required to guess. This is not an intuitively obvious method for measuring anything and can weaken participant confidence in the procedure. For this group yes/no procedures might be more suitable.

This report, therefore, addresses two issues concerning the use of adaptive, single-interval (yes/no) methods for measuring absolute thresholds for tones in quiet. Do they give different results from multiple-interval forced-choice methods and are they more or less efficient? There is already a substantial literature on the statistical aspects of single-interval methods (e.g., Brownlee *et al.*, 1953; Choi, 1990; Cornsweet, 1962; Dixon and Mood, 1948; Levitt, 1971; von Bekesy, 1947) that focuses on appropriate procedures for finding the mean of an underlying psychometric function. However, the subsequent widespread adoption of a signal-detection approach to the nature of threshold (Green and Swets, 1966; Swets, 1964) encouraged the adoption of "objective techniques" such as multiple-interval, forced-choice where the listener's response could be classified as "right" or "wrong." The signal-detection approach seeks to decompose the psychometric function into two components, sensitivity and criterion, claiming that sensitivity is the relevant measure when assessing threshold.

Later, Green (1993) sought to renew interest in the more subjective yes/no approach by emphasizing how efficient it could be. Leek *et al.* (2000), more recently, highlighted the benefits of using this approach, particularly in a clinical context. However, even if it could be shown that adaptive single-interval methods are more efficient, the suspicion remains that subject caution may contaminate the measurements and render them valueless. Kaernbach (1990) previously sought to reconcile the two approaches by introducing trials when no stimulus was presented in a single-interval procedure. This permitted a full assessment of hit rates and false-alarm rates as required by signal-detection theory. This procedure included a sophisticated method for the selection of stimulus levels that promoted substantial efficiencies. Nevertheless, these "subjective" methods remain minority approaches.

The debate has been strongly influenced by Marshall and Jesteadt's (1986) report showing that standard clinical methods for assessing absolute threshold gave overestimates when compared with a two-interval forced-choice (2IFC) methodology. This is often construed as a vindication of the "objective" approach. However, this is a misreading of their results. Marshall and Jesteadt (1986) showed, in the same report, that single-interval methods *when appropriately applied* can yield thresholds similar to those obtained using the 2IFC approach. They attributed the overestimates obtained using the standard clinical methodology to simple, easily remedied, procedural problems.

Marshall and Jesteadt (1986) identified various procedural deficiencies in the clinical approach. The first problem is purely statistical and concerns the clinical practice of choosing the "lowest stimulus level that is reliably heard by the patient." This automatically biases the threshold estimate to be above the 50%-point of the psychometric function by an amount that depends on the step size. A large step size such as the 5-dB step size used in their clinical procedure exaggerates the effect compared to the smaller 2-dB step size used in their comparison 2IFC-procedure. The second factor is psychological and concerns the timing of the presentation of the test stimulus. For 2IFC, the stimulus timing is precisely locked to a visual cue while the clinical procedure has variable timing and no visual cue. When Marshall and Jesteadt (1986) equated these factors using a computer-controlled yes/no procedure, the difference between the estimated thresholds using the two procedures was much smaller. The investigations reported below used precisely timed stimuli with an audible cue and small step sizes in a one-down, one-up procedure in an attempt to minimize the problems identified in their study.

Green (1993) recommended a new method for specifying the sequence of stimulus levels presented to the listener. He suggested that stimulus levels should always be presented at the estimated "sweet point" (most informative level) of the underlying psychometric function, which, in the absence of guessing, would be its 50%-point. This estimate would be updated after each trial by fitting the accumulated data to the best-fit logistic function using a maximum-likelihood (ML) procedure. He found that this procedure was "moderately efficient." This project used Green's (1993) procedure as its starting point. However, it was found that it did not always produce rapid convergence on the true mean and could produce misleading estimates. These flawed estimates had been noticed before (Green, 1995; Leek *et al.*, 2000) but had been identified as secondary consequences of listener lapses of attention. Computer simulations described below, however, show that these errors are intrinsic to Green's (1993) method. As a consequence, it was necessary to evaluate the efficiency of the simpler one-up, one-down rule.

Leek *et al.* (2000) used Green's (1993) method in an extensive study using both normal and clinical populations. Their results indicate that the method is generally acceptable to untrained listeners and gave reliable threshold estimates based on only 24 trials that were comparable with 2IFC-methods. They used catch trials to monitor the false-alarm rate of their listeners although they found that false-alarms were rare (around 5%). The use of catch trials is an import feature of the procedure to be described below where listeners are constrained to keep false-alarms to a minimum and trials containing false-alarms are rejected. While Green (1993) offered a "correction for guessing," an estimation of the guessing rate is possible only after more catch trials than are feasible in practice. This problem will be minimized by keeping rates as close to zero as possible.

The small number of trials used in the study of Leek *et al.* (2000) raises the question of how to estimate the number of trials necessary to achieve a desired precision of threshold estimation. Different studies have different requirements, and it should be possible to adjust the number of trials to take this into account. Computer simulations of the single-interval up/down (SIUD)-procedure using Monte Carlo methods will be used below to give insight into this issue. A simple formula for specifying the required number of trials will be derived on the assumption that the underlying psychometric function takes the form of a logistic curve with a known slope. The results indicate that a surprisingly small number of trials will be necessary in many situations, particularly for hearing impaired listeners with steep psychometric functions.

II. THE SIUD-PROCEDURE

A. Procedure

The SIUD-procedure for measuring absolute thresholds is based on a simple yes/no task. A single stimulus is presented to the participant who responds "yes" or "no" according to whether or not the stimulus was heard. The participant responds by means of a button box linked to a visual display. Part of this display is made invisible when a stimulus is presented thus marking the observation interval. The level of the stimulus is changed from trial to trial using a one-down, one-up adaptive procedure. If the participant says yes, the stimulus level is decreased by a fixed amount. If he says no, the level is increased by the same amount.

The run starts with an initial phase where the stimulus level is set at supra-threshold level (generating a guaranteed yes-response) and is adjusted using a large step size until the first no-response. This initial step size is typically set at 10 dB. The start level is different in each run and randomly located in a range ± 5 dB relative to the nominal start value.

After the first no-response, the stimulus level is set to the mid-point between the previous two levels, and a small step size, say, 2 dB, is used from this point on. The run then continues for a fixed number of trials counting from the trial immediately before the first no-response ("trial 1" in Fig. 1). An illustration of a typical threshold track is shown in Fig. 1.

The choice of a 2-dB small step size was guided by computer simulations (not shown) comparing various step sizes. The 2-dB step gave the lowest variance of the threshold estimates on a range of psychometric slopes (k varied between 0.25 and 1) and trial numbers. In other words, 2-dB steps provide satisfactory reliability in long or short threshold runs. Also, a 2-dB step is commonly used in adaptive

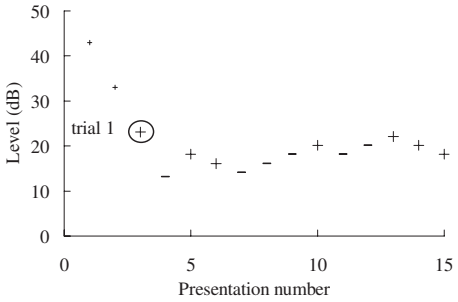


FIG. 1. Illustration of a threshold run. Plus-signs represent the yes-responses, whereas minus-signs represent the no-responses. A large step size of 10 dB is used until the first reversal. After the first reversal the stimulus level is set to the mid-point between the two previous levels and a small step size of 2 dB is used from this point onwards. The trial count starts from the presentation before the first reversal (circled). The responses preceding this point are not included in the threshold estimates (small markers).

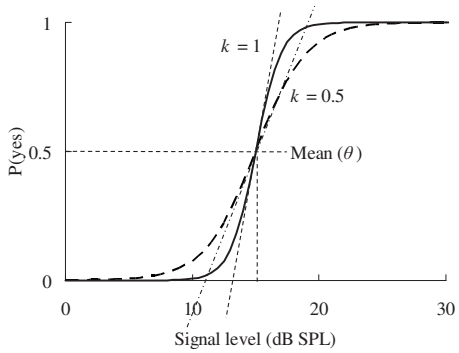


FIG. 2. Psychometric functions with two different slope parameters, $k=1$ (thick continuous line) and $k=0.5$ (thick dashed line). The threshold θ is defined by the mean of the function where the proportion of yes-responses is 0.5.

procedures. Dixon and Mood (1948) suggested a step size close to the standard deviation of the underlying psychometric function.

B. Catch trials

Catch trials are trials where no stimulus is presented and the participant is expected to say no. Catch trials are primarily intended to identify situations where the participant is either not attending or adopting some strategy that is inconsistent with the aims of the investigation. A catch trial is always presented on the second trial in a run to provide a reminder of how "no-stimulus" sounds. 20% of successive trials are catch trials, presented at random without constraint. If the participant is "caught out," the run is stopped and restarted; possibly after resting the participant and giving further instructions. Participants are encouraged not to guess but to report hearing a tone only when they are confident that they have heard it. The restart process following the rare false-alarms acts as an additional incentive for patients to make only confident judgments.

C. Threshold estimation

The threshold is estimated at the end of the run. All stimulus levels from trial 1 (defined above and see Fig. 1) onwards are included in the estimate of the threshold. These are indicated by large markers in Fig. 1. Earlier trials are discarded (small markers in Fig. 1). The threshold can be estimated by using the mean of all these levels (Dixon and Mood, 1948). Alternatively, Cornsweet (1962) suggested that the median level could be used, as this will reduce the effect of any extreme values.

However, in what follows, we are interested in estimating the *accuracy* of our threshold estimate as a function of the number of trials. In this case, it is expedient to estimate the threshold after each trial. We do this by assuming that the participant's decisions close to threshold are approximated by an underlying psychometric function of the form

$$p(L) = 1/(1 + \exp(-k(L - \theta))), \quad (1)$$

where $p(L)$ is the proportion of yes-responses, L is the level of the stimulus [decibel sound pressure level (SPL)], k is a slope parameter, and θ (decibel SPL) is the threshold to be estimated. The threshold θ is the level of the stimulus at which the proportion of yes-responses is 0.5. A psychometric function as described in Eq. (1) is fitted to the responses using a least-squares, best-fit procedure, with θ and k as free parameters.

Figure 2 illustrates this function for two values of slope ($k=0.5$ and $k=1.0$). The value of k typically is close to 0.5 for normal hearing (Green, 1993, using data from Watson *et al.*, 1972). However, Arehart *et al.* (1990) and Carlyon *et al.* (1990) using a d' statistic showed that the slope of the psychometric function can be steeper than normal for patients with a moderate hearing impairment. The slope of the function influences the variability of the threshold estimate. When the slope is steep (continuous line, $k=1.0$), the transition (across level) from yes to no occurs over a narrower range of levels, and fewer trials will be required to estimate the threshold with a given degree of accuracy.

III. EVALUATION I: COMPUTER SIMULATIONS

The accuracy of a threshold estimate improves as the length of a threshold run is increased. The trade-off between accuracy and speed needs to be considered when setting up a measurement protocol, and a compromise is always required. It would therefore be helpful to be able to predict the number of trials needed to obtain a given level of accuracy. In this section "Monte Carlo" computer simulations will be used to assess the improvement in accuracy associated with increasing the number of trials when using the SIUD-procedure. It will be shown that the variability of the estimates can be approximated by a simple mathematical formula and that this formula can be used to specify the number of trials that will be needed to achieve a given level of accuracy.

A. Method

The listener's response was simulated by assuming that it is determined by the psychometric function in Eq. (1). The threshold parameter θ was fixed at 15 dB SPL and the slope value k fixed at 0.5 for the first simulations and at 1.0 for a second evaluation. For each trial, the stimulus level L was used in Eq. (1) to compute the probability p that a yes-response will occur. A uniformly distributed, random number

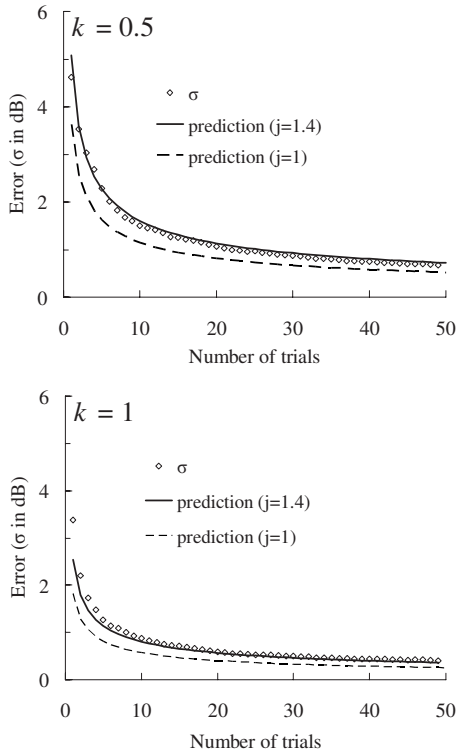


FIG. 3. Standard deviation, σ, of threshold estimates, θ, as a function of the number of trials in a Monte Carlo computer simulation. Top panel: standard deviations (open diamonds) are based on 1000 threshold estimates assuming a psychometric slope k of 0.5. The predictive functions of σ are generated using Eq. (2) with a standard slope value $k=0.5$ and an adjustment factor, $j=1$ (dashed line) or $j=1.4$ (continuous line). Bottom panel: same as top panel but assuming a psychometric slope k of 1. The prediction is generated using Eq. (2) with $k=1$ and $j=1$ (dashed line) or $j=1.4$(continuous line).

between 0 and 1 was then generated to determine the response. If the random number was less than $p(L)$, a yes-response was judged to have occurred on that trial; otherwise, a no-response was assumed. Catch trials were not included in the computer simulations.

The SIUD-procedure was followed exactly as described in Sec. II. The initial starting level was randomly set in a range ± 5 dB relative to a nominal start value of 40 dB SPL. This value was chosen to be above the psychometric function asymptote, guaranteeing only yes-responses on the first presentation. The simulation consisted of 1000 runs. Each run continued for 50 trials. This generated 1000 threshold estimates updated at each of the 50 trial times.

B. Results

The accuracy of the estimates was assessed in terms of the unbiased standard deviations of the thresholds estimated about the true value of θ (i.e., 15 dB SPL). Standard deviations were calculated after each trial across the 1000 runs. The individual data points (open diamonds) in Fig. 3 show how the standard deviation, σ, of the threshold estimate, θ, decreases (i.e., accuracy improves) as the run progresses.

When assuming a slope of 0.5, the accuracy is better than 2 dB after only 10 trials, and after 30 trials, the accuracy is better than 1 dB. The second set of simulations, using a slope parameter $k=1$, shows standard deviations that are lower compared to the standard deviations for a slope of 0.5, and an accuracy of 1 dB is found after less than ten trials. Threshold estimates were approximately normally distributed, and accuracy can be defined to mean that 68.3% of the possible values of the true mean lie within $\pm 1\sigma$ of the true threshold value. No bias in the threshold estimates was observed.

C. Predicting the accuracy of threshold estimates

The formula given in Eq. (1) to describe the hypothesized psychometric function is the logistic function, the cumulative distribution function of the logistic distribution (Hastings and Peacock, 1975). We can therefore use the standard equation for the variance of the logistic probability density function ($\pi^2/3k^2$) to provide an approximate estimate of the reliability of our threshold measurements

$$\sigma = \frac{j\pi}{k\sqrt{3}\sqrt{n}}, \tag{2}$$

where σ is the standard deviation of the threshold estimates, k is the slope parameter of the psychometric function, n is the number of trials in a single threshold run, and j is an adjustment factor to improve the approximation. The fit to the data when $j=1$ is shown in Fig. 3 as a dashed line. It has the correct shape but it underestimates the error.

This underestimation is a consequence of the fact that the stimulus levels are not statistically independent. In our case, each presentation level is related to the previous level by the up/down rule. Some correction is, therefore, required. It is difficult to find a correction factor based on an analytical solution (see Choi, 1990, for a fuller explanation), but a numerical approach based on the Monte Carlo simulations suggests a correction factor j of 1.4. This is illustrated as the continuous line in Fig. 3 for psychometric slopes $k=0.5$ and $k=1.0$. The rms errors of the fit are 0.10 and 0.15 dB, respectively.

D. How many trials?

The number of trials, n, needed to acquire a certain level of accuracy can be calculated by rearranging Eq. (2) as follows:

$$n = \frac{6.4}{k^2\sigma^2}, \tag{3}$$

where σ is the required level of accuracy and a correction factor $j=1.4$ is assumed. Note that the number of trials, n, does not include catch trials or trials in the initial stage (see Fig. 1).

This application of Eq. (2) is considered an important addition to the SIUD-procedure since it allows researchers to predict the number of trials required for a given level of accuracy of the threshold estimates.

In the standard case ($k=0.5$), Eq. (3) reduces to $n=26/\sigma^2$. We can see that in this standard situation a required accuracy (in σ) of 1 dB would indicate the use of 26 trials while an accuracy of 2 dB would indicate a requirement of only 7 trials per threshold run. When the psychometric slope k is 1.0 (for example, for some participants with impaired

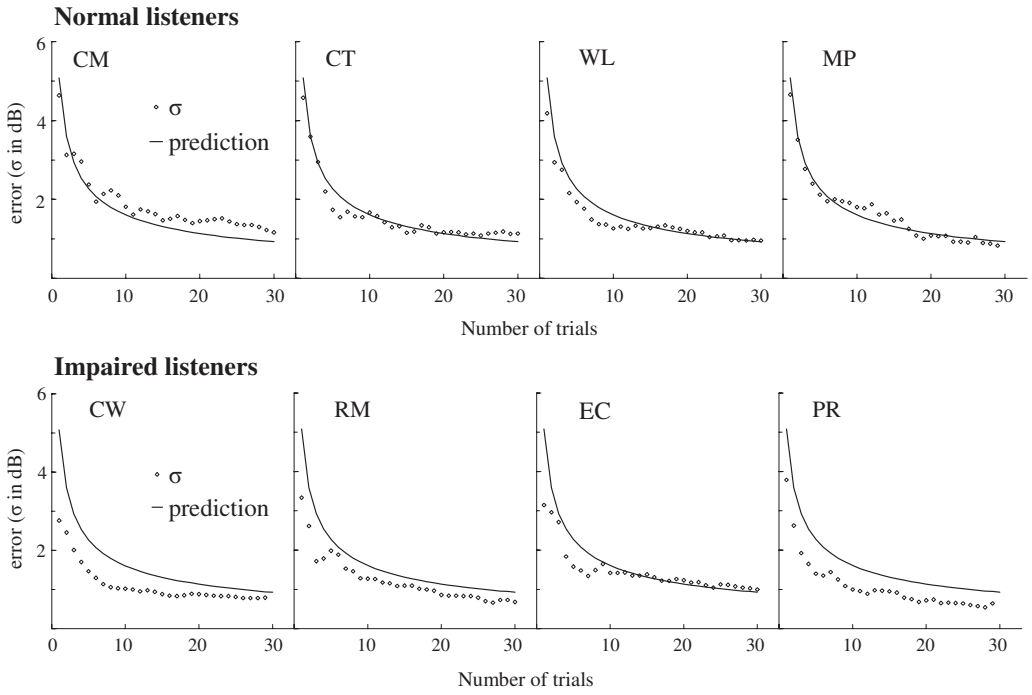


FIG. 4. Standard deviation, σ, of threshold estimates as a function of the number of trials, for four normal listeners (top row) and four impaired listeners (bottom row). Standard deviations (open diamonds) are based on 20 threshold estimates. Continuous lines represent the prediction of the σ values using Eq. (2) with a standard slope value $k=0.5$ and a j-value of 1.4.

hearing), $n=6.4/\sigma^2$ and an accuracy of 1 dB can be achieved with only seven trials. This required number of trials is four times less than for normal listeners.

IV. EVALUTION II: HUMAN LISTENERS

The question remains as to whether these computer simulations of a mathematical abstraction do indeed represent what happens when a human listener is seated in a booth making the same kind of decisions. The predictive value of Eq. (2) was therefore evaluated using human data.

A. Method

Absolute thresholds for a pure tone were measured in four normal and four impaired listeners using the SIUD-procedure described above. For the normal listeners, the stimulus was generally a 2-kHz, 100-ms tone. For one normal listener (MP) the tone frequency was 1 kHz (as a result of an operator error). The participants were aged between 21 and 32 years and have normal audiograms.

The impaired listeners were tested using a 1-kHz, 100-ms stimulus. They were aged between 58 and 76 years. They all had raised thresholds over a large range of frequencies. Participant RM has normal thresholds at the test frequency, but he is considered an impaired listener since he has a sloping loss from 2 kHz onwards. Each participant was tested for 20 threshold runs. Each run consisted of 30 trials. The data were collected in a single session with a 2 min break after every third run. The step size was set at 2 dB and the initial step size was set at 10 dB. Thresholds were estimated after each trial. The adaptive procedure was the same as for the numerical simulations with the addition of 20% catch trials that are not included in the analyses below.

Listeners were seated in a sound-proof booth and stimuli were presented through circumaural headphones (Sennheiser HD600) linked directly to the computer sound card (Audiophile 2496, 24-bit, 96 000-Hz sampling rate). Responses (yes/no) were made using a button box. A monitor in front of the participant showed a display of the button box. While the stimulus was presented displayed button symbols disappeared. Immediately after stimulus presentation the buttons reappeared on the screen signaling that a response was required.

A cue tone at the same frequency and with the same duration as the stimulus tone but 10 dB more intense preceded the stimulus tone by 0.5 s. The cue/stimulus pair was initiated under computer control 0.5 s after the listener's previous response. A raised cosine ramp of 4 ms was applied to both cue and target sounds. When a catch trial occurred, only the cue was presented at the level appropriate if the trial were not a catch trial.

B. Results

The accuracy of the estimates was assessed in terms of the standard deviations, σ, of the threshold estimates θ, after each trial across the 20 runs. These are shown in Fig. 4 for the normal listeners (top row) and for the impaired listeners (bottom row).

The continuous lines in Fig. 4 (top row) show the predicted standard deviations [Eq. (2)] for the normal listeners when assuming a slope parameter $k=0.5$ and applying the correction factor $j=1.4$ (see above). The predictive function fits the normal data with an average rms error across listeners of 0.24 dB.

The standard deviations found in the impaired group

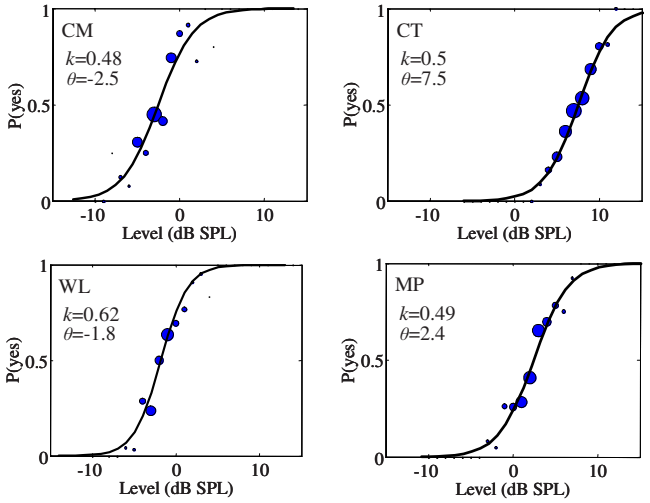


FIG. 5. (Color online) Psychometric function for 4 normal listeners based on 600 yes-/no-responses. The size of the circles represents the number of responses contributing to that point of the psychometric function. The full line represents the best-fit logistic function to the responses. The inset shows the slope parameter k and threshold θ (dB SPL) associated with the best fit. The stimulus was a 100-ms tone with tone frequency 2 kHz for listeners CM, CT, and WL, and 1 kHz for listener MP.

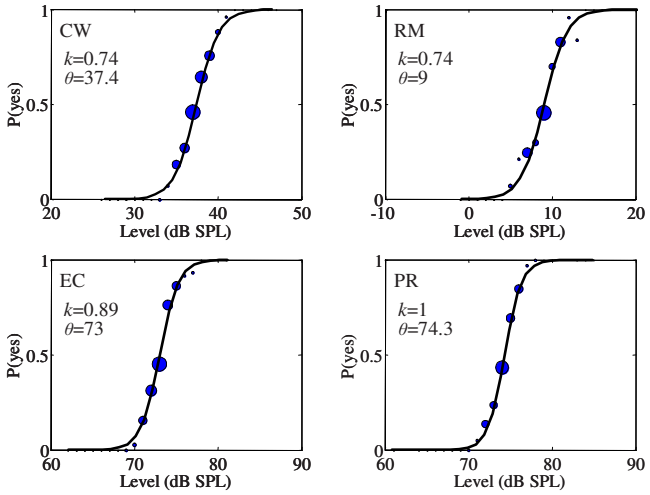


FIG. 6. (Color online) Psychometric functions for four impaired listeners (see Fig. 5 for more details). The stimulus was a 100-ms tone with tone frequency 1 kHz.

were lower than the normal group (Fig. 4, bottom row). The predictive function used for the normal listeners (assuming a slope parameter $k=0.5$ and a correction factor j of 1.4) was also fitted to the data of the impaired listeners. In almost all cases this results in a conservative prediction of the error of the threshold estimates, and this is consistent with the possibility that the impaired listeners have steeper psychometric functions ($k>0.5$).

C. Conclusion

Equation (2) offers a general guide to the accuracy of the SIUD-method as a function of number of trials in a threshold run, particularly if the slope parameter of the psychometric function k is known. If the slope is not known, then a slope parameter $k=0.5$ (and correction factor $j=1.4$) gives a useful, if sometimes conservative, estimate. Equation (3) can be used to decide how many trials are needed to achieve a required level of accuracy.

V. PSYCHOMETRIC SLOPES OF NORMAL AND IMPAIRED LISTENERS

The computer simulations described in Sec. III assumed that the psychometric slope k was either 0.5 or 1. To check these assumptions, the behavioral data collected in the previous experiment (Sec. IV) were reanalyzed to establish appropriate values for the slope.

A. Method

The 600 yes-/no-responses at various signal levels, obtained when measuring the thresholds described in the previous experiment (Sec. IV), were used to generate a psychometric function for each normal and impaired listener (Figs. 5 and 6). Responses were aggregated into bins of 1-dB width, and the proportion of yes-responses in each bin was calculated. The relative size of the circles is used to indicate the relative number of responses at each stimulus level. The best-fit function [Eq. (1)] is shown as the continuous line through the data points. The k-value and threshold θ, associated with this best fit, are shown in the insets.

B. Results

Figure 5 shows the psychometric functions of the four normal listeners. The slope estimates k range from 0.48 for listener CM to 0.62 for listener WL. The psychometric functions for the four impaired listeners are shown in Fig. 6. Their slopes (0.74, 0.74, 0.89, and 1.00) were considerably steeper. The observation of steeper slopes is in line with other studies (Arehart *et al.*, 1990; Carlyon *et al.*, 1990).

VI. COMPARISON WITH OTHER PROCEDURES

The use of the SIUD-method can only be recommended if it is as accurate as other methods currently used in research laboratories. The SIUD-procedure was therefore compared with two other procedures in common use: (1) 2IFC, two-down/one-up method described by Levitt (1971) and (2) the single-interval, ML-method of Green (1993).

A. Computer simulations

1. Method

Monte Carlo simulations were made in the same manner as described above. Numerical simulations assumed an underlying psychometric function given in Eq. (1) where the true threshold θ was fixed at 15 dB SPL, and the slope of the psychometric function k was fixed at 0.5. Threshold estimates were simulated over 500 runs for all three procedures. The step size was always 2 dB except for the initial step size (10 dB).

In all conditions, the initial starting level was randomly set in a range ± 5 dB relative to a nominal start value of 40 dB SPL. Catch trials were not used in the SIUD-condition. A threshold estimate was computed at the end of each run and the standard deviation computed over the 500 runs.

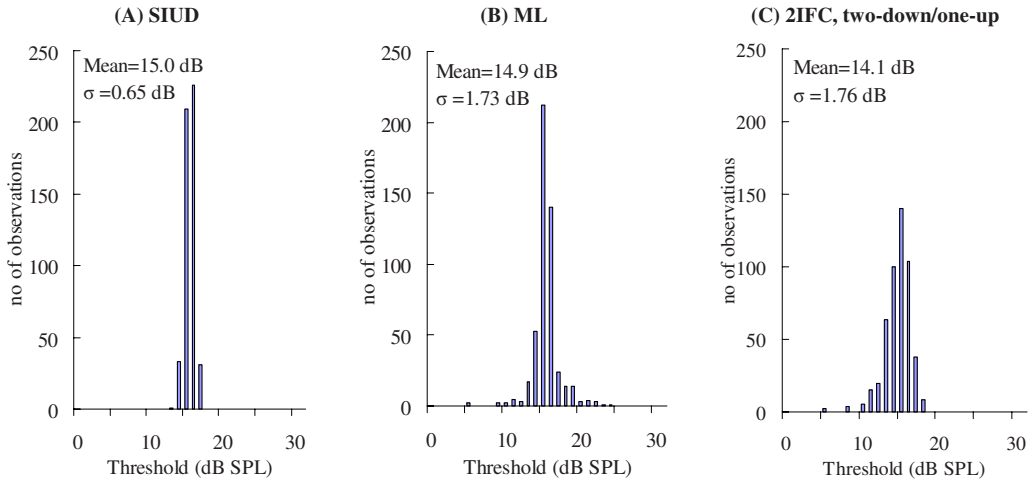


FIG. 7. (Color online) Histogram for threshold estimates using (a) SIUD, (b) ML, and (c) 2IFC, two-down/one-up. 500 threshold estimates were obtained per condition. Threshold runs were terminated after 50 trials for the ML- and SIUD-condition (not including the initial trials) and after 14 reversals in the 2IFC-condition. The inset shows the mean threshold estimate and the standard deviations (σ) for each condition.

(a) *2IFC.* The thresholds for the 2IFC-procedure were measured using the two-down/one-up adaptive procedure described by Levitt (1971). The signal level was initially adjusted in 10-dB steps. After two reversals, the step size was reduced to 2 dB. Each run was terminated after 14 reversals. The threshold was computed by averaging the levels of the last 12 reversals. The 2IFC-condition was the first to be simulated. It was found that the average number of trials required for 14 reversals was 50. The other two conditions were then simulated using this number of trials so that a fair comparison of accuracy could be made.

(b) *ML.* The ML-procedure followed as closely as possible the protocol described by Green (1993). The initial step size (10 dB) was used to adjust the stimulus level up to the first reversal. After that, the ML-procedure was used to set the new stimulus level after each trial. The best-fit psychometric function [see Eq. (1)] was obtained using the ML-method described by Green (1993) on the basis of all observations up to that point. The 50%-point of the function was then used to determine the level of the next stimulus to be presented. Each run continued for 50 trials (not including the initial trials). The threshold estimate for the run was taken to be the 50%-point of the final best-fit psychometric function. The false-alarm rate was fixed at zero.

(c) *SIUD.* The SIUD-procedure was exactly as described above using 50 trials per threshold run.

2. Results

The distribution of the threshold estimates across 500 runs for all three procedures (SIUD, ML, and 2IFC) is given in Fig. 7. The mean threshold estimates are 15.0, 14.9, and 14.1 dB SPL for SIUD, ML, and 2IFC, respectively. The mean standard deviations σ are 0.65, 1.73, and 1.76 dB for SIUD, ML, and 2IFC, respectively.

The threshold estimate for the 2IFC simulation (14.1 dB SPL) is below the true threshold (15 dB). This is partly because the 2IFC-procedure estimates the 70.7%-point of a psychometric function where the minimum hit rate is 50%. The equivalent point on the true psychometric function (ranging from 0% to 100% correct) is 41.4%, i.e., an underestimate of the true mean (see Fig. 8). For a slope k of 0.5, an adjustment of +0.7 dB is necessary to establish the level at the 50%-point of the underlying (yes/no) psychometric function. This adjustment was previously suggested by Leek *et al.* (2000). The new threshold estimate of 14.8 dB SPL is closer to, but still an underestimate of, the true threshold.

Our main concern here, however, is *reliability* as represented by the standard deviation of the threshold estimates over many runs. The spread of threshold estimates in the SIUD-condition is considerably less than the spread in the ML-condition or the 2IFC-condition.

The lower accuracy of the ML-procedure is, at least partly, explained by a number of extreme threshold estimates both above and below the true threshold. These can be seen as unexpected outliers in the distribution of threshold estimates in Fig. 7(B). To investigate the matter further, a limited set of tracks of the threshold estimates is considered

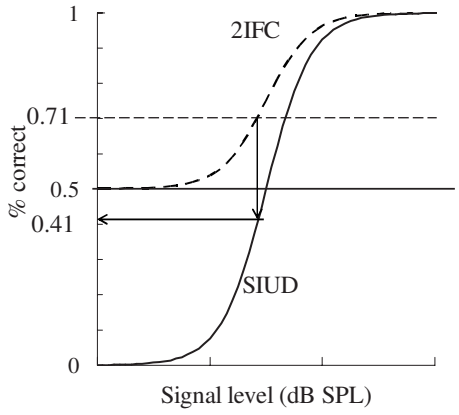


FIG. 8. Schematic representation of the psychometric function for a 2IFC-condition (dashed line) ranging from 50% to 100% correct responses and the psychometric function generated in a single-interval procedure (continuous line) ranging from 0% to 100% correct responses. The 70.7%-point on the 2IFC-psychometric function corresponds to the 41.4%-point on the SIUD-psychometric function.

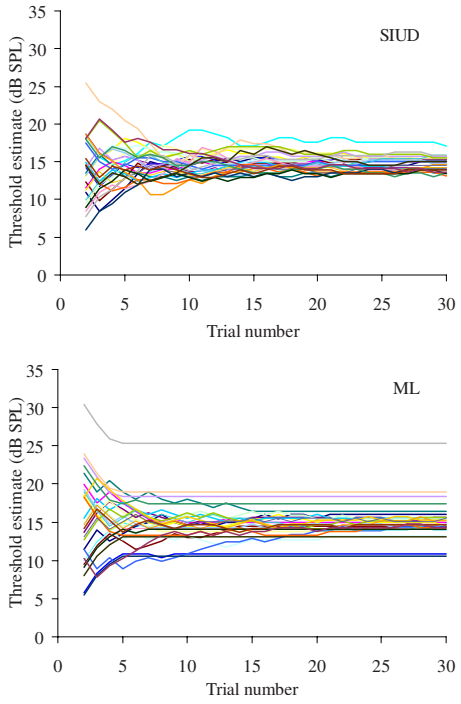


FIG. 9. (Color online) Threshold estimate as a function of trial number in a SIUD-procedure (top panel) and a ML-procedure (bottom panel) obtained using Monte Carlo simulations. Threshold tracks associated with 30 threshold runs are pictured for each condition. Each run consists of 30 trials.

individually. These are shown in Fig. 9. After each observation, the ML-procedure estimates the 50%-point of the psychometric function. The bottom panel of Fig. 9 shows that these estimates normally converge quickly on a value close to the true mean (15 dB) of the function. However, some tracks settle quickly and permanently on a *false estimate*. These rogue estimates inflate the overall standard deviation of the threshold estimation procedure.

The top panel of Fig. 9 shows equivalent tracks for the SIUD-condition. In this case the tracks all appear to be converging on the true threshold. This agrees with Fig. 7(A) where no outliers in the distribution are present.

The relatively widespread of estimates in the 2IFC-procedure may also be partly explained by outliers. For example, the distribution of estimates in Fig. 7(C) shows two very low estimates. Moreover, the distribution of estimates is asymmetric with the long tail toward values below the mean. This is almost certainly attributable to random adjustments in level when the stimulus is below threshold and the patient is required to guess. During this guessing phase, the patient may guess correctly and the stimulus level reduces even further below threshold. At this time there is a chance that a random walk be initiated with peaks and troughs below threshold levels. In this case the estimated threshold will be an underestimate of the true threshold. A presentation of the individual threshold tracks in the 2IFC-condition is not included in Fig. 9 since the nature of this procedure does not allow for threshold estimates to be made after each trial.

B. Human data

As a final reassurance concerning the reliability of the SIUD-procedure, we compared the thresholds measured using both SIUD and 2IFC of a group of student volunteers with no prior experience of audiometric methods.

TABLE I. Individual thresholds (in dB SPL) and standard deviations, σ, for the SIUD and 2IFC-condition. In the 2IFC-condition the adjusted thresholds are between parentheses (targeting the 50%-point on a psychometric function ranging from 0% to 100% correct). The bottom row shows the average thresholds and average standard deviations across listeners for each condition.

	SIUD		2IFC	
Listener	Threshold	σ	Threshold	σ
S1	16.6	2.0	16.3 (17.0)	4.6
S2	8.2	1.9	6.4 (7.1)	1.7
S3	10.9	1.8	10.0 (10.7)	1.9
S4	−2.4	2.3	−1.1 (−0.4)	2.2
S5	11.7	1.9	5.0 (5.7)	6.7
S6	5.4	1.3	3.1 (3.8)	3.1
S7	9.0	2.2	8.4 (9.1)	1.6
S8	8.9	1.9	9.2 (9.9)	4.3
S9	5.2	0.7	4.2 (4.9)	4.4
Average	8.2	1.8	6.9 (7.6)	3.4

1. Method

Nine listeners were used with audiometric thresholds within the normal range. The procedures used are exactly as described above. Five threshold estimations were made using both the SIUD and 2IFC-procedure. The SIUD-condition used only ten trials per run. The 2IFC used eight reversals and the thresholds were estimated as the mean of the last six reversal levels. This required an average of 32 trials per run.

2. Results

Table I shows the average threshold and associated standard deviation for each condition for each individual listener. The average threshold across all listeners for the SIUD-method was 8.2 dB SPL. The average for the 2IFC-thresholds was 6.9 dB SPL. After applying the 50%-adjustment suggested above (Sec. VI A) the mean 2IFC-threshold (across listeners) is 7.6 dB SPL. These adjustments are shown between parentheses in Table I alongside the initial 2IFC-thresholds.

The average thresholds per listener are similar for both conditions. Although the average threshold across listeners is slightly higher for the SIUD-condition compared to the 2IFC-condition, this was the case for only five out of nine listeners. This suggests that there is no consistent pattern for SIUD-threshold to be higher than 2IFC-thresholds. The standard deviation of the threshold estimates, however, is similar or substantially higher for the 2IFC-condition (average $\sigma = 3.4$) compared to the SIUD-condition (average $\sigma = 1.8$). These findings are consistent with the Monte Carlo simulations presented above (Sec. VI A). The average standard deviation across listeners in the SIUD-condition is half the standard deviation in the 2IFC-condition despite the fact that almost three times fewer trials were used in the SIUD-condition.

The average number of catch trials presented per threshold run in the SIUD-condition was 3.6. A caught-out incident occurs when a listener reports to have heard the stimulus when no stimulus was presented. The rate of caught-out incidents had an overall average of 2.1% of the catch trials. Six of the nine participants had a zero caught-out rate. The average in the remaining three listeners (S2, S4, and S9) was 6.9%.

C. Conclusion

The SIUD-method produces a narrower spread of threshold estimates than either the ML- or the 2IFC-method over runs of 50 trials, and shows no obvious tendency to overestimate the threshold.

VII. DISCUSSION

In summary, the SIUD-procedure is recommended as a fast and reliable threshold procedure to estimate absolute threshold in both normal and impaired groups of listeners. Numerical simulations suggest that it is substantially more efficient than either Green's (1993) ML-method or the conventional 2IFC-procedure. They also showed that the number of trials needed to estimate threshold using the SIUD-method can be specified approximately using a simple formula based on the required accuracy and the steepness of the psychometric slope. Psychometric slopes were found to be steeper for hearing impaired listeners and, as a consequence, fewer trials are required for this group.

The new results reported here extend Leek *et al.*'s (2000) and Green's (1993) studies of single-interval methods by allowing the number of trials to be varied according to the accuracy requirements of the study. For example, if an accuracy of ± 2 dB is adequate then only seven trials are needed resulting in a considerable saving in testing time over other laboratory practices. Participants with a hearing impairment will often have steeper psychometric slopes than normal hearers. In this case even fewer trials will be needed. A fourfold reduction in the required number of trials applies if the slope is as steep as 1.0. In our experience with impaired hearers, ten trials give thresholds that are satisfactory for a repeatable clinical description of the impairment.

The single-interval method has been discussed solely in the context of absolute thresholds. Clearly, it could also be used in the context of a wide range of threshold measurements. It must be stressed, however, that the estimate of the number of required trials must be based on knowledge of the slope of the underlying psychometric function. If suprathreshold levels are used, compression may apply and the slope will be more shallow (Schairer *et al.*, 2008). This implies that more trials will be required.

Green's (1993) ML-method was found to be subject to occasional false estimates that arise from time to time as the result of a premature convergence on an inappropriate threshold value. These erroneous values make a proper comparison of efficiency inadmissible. They were noted by Leek *et al.* (2000) as well as by Green (1995) where they were attributed to secondary consequences of "attentional lapses." However, the numerical simulations showed that they are intrinsic to the procedure itself. Moreover, the simulations also show that these false estimates do not improve if the number of trials is increased (Fig. 9, bottom panel). In Green's (1993) procedure each successive stimulus level is set at the current best estimate of the threshold. Unfortunately, this is self-reinforcing and a false estimate can quickly become permanent across an indefinite number of trials. In contrast, the SIUD-method is not subject to the same problem and will always eventually converge on the true threshold.

Further simulations and experimental observations showed that single-interval methods gave similar threshold estimates to the 2IFC approach. This result replicates the findings of Leek *et al.* (2000) as well as Marshall and Jesteadt (1986). While the latter study was primarily aimed at comparing 2IFC with standard clinical practice based on the ANSI (1997) threshold search strategy, they did also include a single-interval comparison condition using the method of constant stimuli. Their single-interval procedure gave similar results to 2IFC.

Adherents of the signal-detection theory of the nature of absolute threshold will be puzzled by our finding that SIUD- and 2IFC-thresholds obtained using human listeners did not show any systematic differences. Of course, the absence of any measured effect does not prove that none exists; one might be observed in different testing circumstances where listeners choose to apply extra caution to their judgments. However, our listeners were asked to be very cautious and no difference was seen. The matter clearly invites further investigation. If response bias is a matter of concern, however, Kaernbach's (1990) single-interval adjustment-matrix (SIAM) procedure uses single-interval methodology while taking listener's criterion into account.

Gu and Green (1994) recommended that catch trials be used to estimate the listener's "guessing rate." However, we abandoned this approach because it was impossible to obtain an accurate estimate of guessing based on only a small number of catch trials. We were, however, reassured by the estimates of Leek *et al.* (2000) who found very low guessing rates. We encouraged our listeners to be conservative in their judgments. We defended this to them on the grounds that it was impossible to make useful measurements if listeners reported yes when no stimulus had been presented. In the event, our listeners gave very few "false-alarms." When they did occur, the run was restarted and this further discouraged guessing. Catch trials are, however, a useful guide to the attentional state of the listener and rest periods can be arranged if they begin to occur during a measurement session. Moreover, catch trials offer a regular reminder to the listener of what a no-stimulus condition sounds like. This may add to listener's confidence later when a stimulus is presented just above threshold.

Marshall and Jesteadt (1986) drew attention to the importance of the visual cue normally given in 2IFC-procedures to help the listener pay attention at the right time. This effect had previously been studied by Watson and Nichols (1976) who found a 2-dB improvement in threshold when an appropriate cue was given. The procedure followed in this study involved giving an audible cue 0.5 s before the

test stimulus. The cue has the same frequency and the same duration as the test stimulus and may well have minimized errors due to temporal uncertainty.

VIII. CONCLUSIONS

We recommend that the SIUD-method with catch trials be used for studies where it is necessary to limit the number of trials as much as possible. The procedure is simple to administer and requires little participant training. The estimated thresholds are comparable in value and are less variable than commonly used ML- and 2IFC-method. They also yield a known degree of accuracy for a given number of trials. The use of an audible cue similar to, but preceding, the test stimulus by a fixed time interval is also recommended as an aid to better threshold estimation.

ACKNOWLEDGMENTS

The authors would like to thank Christine Tan and Soumini Menon for assisting in data collection. During the course of this study, the authors benefited from helpful discussions with Marjorie Leek concerning procedural issues and Graham Upton concerning statistical issues. Helpful comments on a preliminary draft manuscript were also received from Graham Upton, Enrique Lopez-Poveda, and Brian Walden. They would also like to thank the two anonymous reviewers who made insightful comments on a previous version of this article.

ANSI S3.21-1978 (**1997**). American National Standard Method for Manual Pure-Tone Threshold Audiometry (American National Standards Institute, New York).

Arehart, K. H., Burns, E. M., and Schlauch, R. S. (**1990**). "A comparison of psychometric functions for detection in normal-hearing and hearing-impaired listeners," J. Speech Hear. Res. **33**, 433–439.

Brownlee, K. A., Hodges, J. L., and Rosenblatt, M. (**1953**). "The up-and-down method with small samples," J. Am. Stat. Assoc. **48**, 262–277.

Carhart, R., and Jerger, J. F. (**1959**). "Preferred method for clinical determination of pure-tone thresholds," J. Speech Hear. Disord. **24**, 330–345.

Carlyon, R. P., Buus, S., and Florentine, M. (**1990**). "Temporal integration of trains of tone pulses by normal and by cochlearly impaired listeners," J. Acoust. Soc. Am. **87**, 260–268.

Choi, S. C. (**1990**). "Interval estimation of the LD50 based on an up-and-down experiment," Biometrics **46**, 485–492.

Cornsweet, T. N. (**1962**). "The staircase-method in psychophysics," Am. J. Psychol. **75**, 485–491.

Dixon, W. J., and Mood, A. M. (**1948**). "A method for obtaining and analyzing sensitivity data," J. Am. Stat. Assoc. **43**, 109–126.

Green, D. M. (**1993**). "A maximum-likelihood method for estimating thresholds in a yes-no task," J. Acoust. Soc. Am. **93**, 2096–2105.

Green, D. M. (**1995**). "Maximum-likelihood procedures and the inattentive observer," J. Acoust. Soc. Am. **97**, 3749–3760.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Gu, X., and Green, D. M. (**1994**). "Further studies of a maximum-likelihood yes-no procedure," J. Acoust. Soc. Am. **96**, 93–101.

Hastings, N. A. J., and Peacock, J. B. (**1975**). *Statistical Distributions: A Handbook for Students and Practitioners* (Butterworth, London).

Kaernbach, C. (**1990**). "A single-interval adjustment-matrix (SIAM) procedure for unbiased adaptive testing," J. Acoust. Soc. Am. **88**, 2645–2655.

Leek, M. R., Dubno, J. R., He, N., and Ahlstrom, J. B. (**2000**). "Experience with a yes-no single-interval maximum-likelihood procedure," J. Acoust. Soc. Am. **107**, 2674–2684.

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

Marshall, L., and Jesteadt, W. (**1986**). "Comparison of pure-tone audibility thresholds obtained with audiological and two-interval forced-choice procedures," J. Speech Hear. Res. **29**, 82–91.

Schairer, K., Messersmith, J., and Jesteadt, W. (**2008**). "Use of psychometric-function slopes for forward-masked tones to investigate cochlear nonlinearity," J. Acoust. Soc. Am. **124**, 2196–2215.

Swets, J. A. (**1964**). *Signal Detection and Recognition by Human Observers* (Wiley, New York).

von Bekesy, G. (**1947**). "A new audiometer," Acta Oto-Laryngol. **35**, 411–422.

Watson, C. S., Franks, J. R., and Hood, D. C. (**1972**). "Detection of tones in the absence of external masking noise. I. Effects of signal intensity and signal frequency," J. Acoust. Soc. Am. **52**, 633–643.

Watson, C. S., and Nichols, T. L. (**1976**). "Detectability of auditory signals presented without defined observation intervals," J. Acoust. Soc. Am. **59**, 655–668.

Matching the waveform and the temporal window in the creation of experimental signals

William M. Hartmann[a] and Eric M. Wolf
Department of Physics and Astronomy, Michigan State University, 1226 BPS Building, East Lansing, Michigan 48824

(Received 24 October 2008; revised 29 July 2009; accepted 31 July 2009)

When a periodic waveform with a discrete-harmonic spectrum is temporally windowed to make a signal, its spectrum becomes a continuous function of frequency. However, there are discrete-frequency representations for windowed signals such as the Fourier series representation of a periodically extended signal. This article introduces the concept of matching between the temporal window and the periodic waveform. Matching leads to a discrete-frequency representation in which the Fourier transform of the windowed signal preserves the amplitudes and phases of the waveform on the set of original waveform frequencies. Generating signals with matched window and waveform leads to important control of experiments.
 [DOI: 10.1121/1.3212928]

PACS number(s): 43.66.Yw, 43.64.Yp, 43.66.Pn [MW] Pages: 2580–2588

I. INTRODUCTION

Auditory scientists generate experimental stimuli, signals, or noise, for presentation to humans and other animals. Often those stimuli are specified in terms of their spectral properties. Psychoacoustician A might specify a broadband complex tone having harmonics with equal amplitudes and Schroeder phases. Physiologist B might specify a narrowband noise with Rayleigh-distributed amplitudes and alternating phases. Hearing scientist C might specify a five-component signal with constant phases. In all these cases, the experimenter expects that the specified spectral properties will be preserved, at least to a good approximation, in the stimulus presented to the listener.

A. Spectral properties

When a stimulus is initially defined in terms of exact discrete spectral properties, it can be represented as a sum of cosines with component amplitudes C_n,

$$x(t) = \sum_{n=0}^{N} C_n \cos(\omega_n t + \phi_n), \tag{1}$$

where component angular frequencies ω_n and phases ϕ_n are arbitrary. In what follows, this wave will be called the "unlimited-duration waveform" or the "waveform." Because its duration is unlimited, it has a bandlimited Fourier transform given by

$$X(\omega) = \int_{-\infty}^{\infty} dt\, e^{-i\omega t} x(t) \tag{2}$$

or

$$X(\omega) = \pi \sum_{n=-N}^{N} X_n \delta(\omega - \omega_n), \tag{3}$$

where $X_n = C_n \exp(i\phi_n)$, $\omega_{-n} = -\omega_n$, $C_{-n} = C_n$, and $\phi_{-n} = -\phi_n$. With these conditions, X_{-n} is equal to the complex conjugate, X_n^*, and $x(t)$ is a real function (see, for instance, Hartmann, 1998).

B. Finite-duration signals

The infinitely sharp spectrum of Eq. (3) applies only to an unlimited-duration waveform. In practice, a signal is presented to a listener with a finite total duration, T_D. The finite-duration signal is created by multiplying waveform $x(t)$ by a temporal window function, $w(t)$, to create the final signal $y(t)$,

$$y(t) = w(t)x(t). \tag{4}$$

The finite duration leads to the well-known "spectral splatter" wherein the power spectrum acquires power outside the frequency band of the original waveform. The spectrum becomes a continuum with all frequencies, inside and outside the waveform band, represented more or less.

We consider the case in which signal $y(t)$ can be represented on the finite interval by a Fourier series,

$$y'(t) = \sum_{n=1}^{N'} C'_n \cos(2\pi n t/T_D + \phi'_n) \quad (0 < t \le T_D), \tag{5}$$

where the designation $y'(t)$ distinguishes the finite-duration signal from $y(t)$, which is defined for all time. The fundamental angular frequency of the series is $2\pi/T_D$, a function of the window duration. All the other frequencies in the series are harmonics, $n(2\pi/T_D)$. The finite-duration signal can be periodically extended forward and backward in time to fill the entire time axis, creating the periodically-extended signal,

[a]Author to whom correspondence should be addressed. Electronic mail: hartmann@pa.msu.edu

$$\underline{y}(t) = \sum_{n=1}^{N'} C'_n \cos(2\pi nt/T_D + \phi'_n) \quad (-\infty < t < \infty). \tag{6}$$

At this point in the development, we introduce the concept of matching the temporal window and the waveform. In the simplest case, the temporal window that defines the finite interval in Eq. (5) is rectangular. If the original unlimited-duration waveform $x(t)$ contains power only on a set of harmonic frequencies $\{\omega_n = n\omega_o\}$ where the fundamental angular frequency ω_o happens to be related to the duration T_D by $\omega_o = 2\pi/T_D$, then the terms in the Fourier series $y'(t)$ are the same as the terms in the sum for the waveform $x(t)$. Specifically, $C'_n = C_n$, $\phi'_n = \phi_n$, and $N = N'$. The window and the waveform are matched. Because the duration T_D is equal to the period of the waveform, $x(t)$, the periodically-extended signal $\underline{y}(t)$ is the same as $x(t)$.

Equation (5), in terms of the duration T_D, is not the only possible Fourier series. It is possible to represent y on the finite interval T_D in terms of any long time, T_L,

$$y_L(t) = \sum_{n=1}^{N'} C_{L,n} \cos(2\pi nt/T_L + \phi_{L,n}), \tag{7}$$

so long as $T_L > T_D$. However, to do that, the amplitudes and phases, $C_{L,n}$ and $\phi_{L,n}$, must be chosen to force $y_L(t)$ to be zero within the part of the T_L interval that is not included in the T_D interval. The frequencies, amplitudes, and phases in this representation do not agree with those in the original periodic waveform, $x(t)$.

C. Temporal windows

The section above, unifying the waveform, the Fourier series, and the periodically-extended signal when the waveform and window are matched assumes a rectangular temporal window. Although the rectangular window leads to simple mathematics, it is not often used in practice. A rectangular window produces discontinuities in the signal and/or its derivatives at the onset and offset; these cause audible clicks that may be distracting to human or animal listeners. The clicks by themselves may be spurious stimuli, detracting from the intended purpose of the band-limited amplitude and phase cues of the desired stimulus. In order to reduce onset and offset clicks, it is usual to apply a temporal window which turns the stimulus on and off more gradually.

When a window other than a rectangular window is applied to the waveform, interesting possibilities arise. It is possible to retain the periodically-extended signal, which includes information about the temporal window. Alternatively, and this is the point of the present article, it is possible to maintain the matching concept so that the Fourier transform of the windowed signal preserves the spectral amplitudes and phases of the waveform on the set of waveform frequencies. If the window and waveform are not matched, the spectrum is not preserved on any set of frequencies. Then the spectrum of the signal presented to the listener becomes out of control, more or less depending on details. It seems likely that experimenters often use temporal windows and waveforms that are *not matched* because the matching conditions are not immediately obvious, as described below.

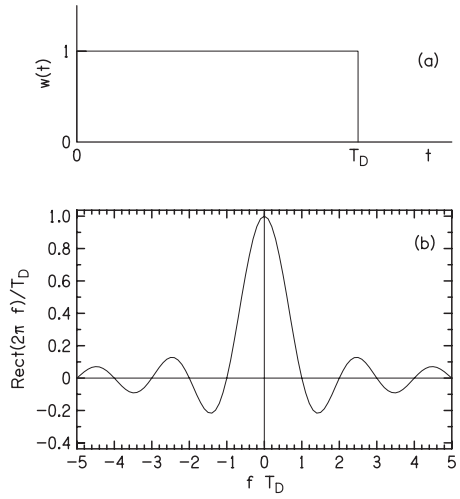


FIG. 1. Rectangular temporal window. (a) The window as a function of time. (b) The Fourier transform of the rectangular window apart from the phase factor. The Fourier transform is zero for integer values of fT_D.

II. SPECTRAL CONSEQUENCES OF TEMPORAL WINDOWING

The spectral consequences of temporal windowing appear in the Fourier transform of $y(t)$, namely, $Y(\omega)$. Because $y(t) = w(t)x(t)$ is a product, $Y(\omega)$ is given by the convolution

$$Y(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega' W(\omega - \omega')X(\omega'). \tag{8}$$

Because $X(\omega)$ is a sum of delta functions from Eq. (3), the integral is easy to do, and

$$Y(\omega) = \frac{1}{2}\sum_n X_n W(\omega - \omega_n). \tag{9}$$

According to Eq. (9) the spectrum of the final windowed signal can be found from the spectrum of the infinite-duration waveform if we know the Fourier transform of the temporal window, $W(\omega)$.

A. Rectangular window

The rectangular window extends from $t=0$ to $t=T_D$, as shown in Fig. 1(a). The Fourier transform is W_{Rect}, given by

$$W_{\text{Rect}}(\omega) = \text{Rect}(\omega)e^{-i\omega T_D/2}, \tag{10}$$

where "Rect(ω)" is

$$\text{Rect}(\omega) = \frac{2\sin(\omega T_D/2)}{\omega}. \tag{11}$$

This Fourier transform is the product of two factors. Function Rect is the Fourier transform of the rectangular window translated backward in time by half its duration so that it is symmetrical about the origin. Because of this symmetry, Rect is entirely real. The other factor is a phase factor, $\exp(-i\omega T_D/2)$, and its role is to translate the rectangle forward again, back to where it belongs.

Function $\mathrm{Rect}(2\pi f)/T_D$ is plotted in Fig. 1(b). $\mathrm{Rect}(2\pi f)/T_D$ is a sinc function. The figure shows that the sinc function is zero for integer values of fT_D. The duration of time that appears in the argument of the sinc function (T_D for the rectangular window) will be called the "significant duration, T_S."

Knowing the Fourier transform of the window, we now know the Fourier transform Y,

$$Y(\omega) = \frac{T_D}{2}\sum_n X_n \frac{\sin[(\omega-\omega_n)T_D/2]}{(\omega-\omega_n)T_D/2} \times \exp[-i(\omega-\omega_n)T_D/2]. \quad (12)$$

Transform $Y(\omega)$ includes all the details of the temporally windowed signal $y(t)$. If the waveform and the rectangular window are matched, as described in Sec. I, then waveform frequencies ω_n are given by $\omega_n = n\omega_o$, where $\omega_o T_D = 2\pi$. When $Y(\omega)$ is evaluated at those frequency values, we find a revealing equation,

$$Y(m\omega_o) = \frac{T_D}{2}\sum_n X_n \frac{\sin\,\pi(m-n)}{\pi(m-n)} \exp[-i\pi(m-n)]. \quad (13)$$

The sinc function in the sum is zero for all values of n except when $n=m$. When $n=m$ then the sinc function equals 1. Thus the sinc function has become a Kronecker delta function, and Eq. (13) simplifies to

$$Y(m\omega_o) = \frac{T_D}{2} X_m. \quad (14)$$

Because Eq. (14) holds for both positive and negative values of m, both the amplitude and the phase of the original, unlimited-duration signal are correctly represented in the Fourier transform $Y(\omega)$ when evaluated at the special frequencies $\omega = m\omega_o$. In this sense, the spectrum is preserved. The values of the Fourier transform at these special frequencies $Y(m\omega_o)$ are related to the Fourier series coefficients in Eq. (5) by a constant factor,

$$Y(m\omega_o) = \frac{T_D}{2} C'_m \exp(i\phi'_m). \quad (15)$$

The development of this last equation gives insight as to why the spectrum is preserved on the set of frequencies $\{\omega_m = m\omega_o\}$. The reason is that the unlimited-duration waveform is matched to the rectangular window because its fundamental frequency ω_o is equal to $2\pi/T_D$ causing the sinc function to be a Kronecker delta function. Matching has effectively made the window disappear for these special frequencies. For the rectangular window, the total duration T_D is also the significant duration. As an example, if we would like to use a 100-ms, rectangularly-windowed signal, we would choose the frequencies to be integer multiples of 10 Hz, i.e., $\omega_0 = 2\pi \cdot 10$, to obtain a complete set of expansion functions. In Secs. II B and II C, it will be shown that the same principle can be used to match the waveform (characterized by a fundamental ω_o) and the time window (characterized by a significant duration, T_S) for other forms of time window. An interesting alternative window is the raised cosine.

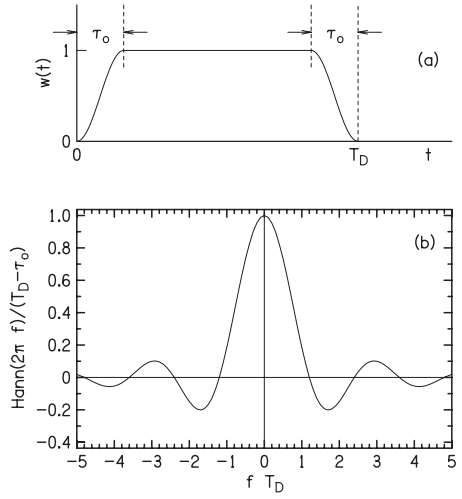


FIG. 2. Raised-cosine temporal window for $\tau_o = T_D/6$. (a) The window as a function of time. The full-on duration is $T_D - 2\tau_o$. (b) The Fourier transform of the raised-cosine window apart from the phase factor. The Fourier transform is zero when fT_D is somewhat larger than integer values.

B. Raised-cosine window

The raised-cosine window (derived from the Hanning or Hann window) is shown in Fig. 2(a). It has a onset edge given by $[1-\cos(\pi t/\tau_o)]/2$, where τ_o is the edge duration. The overall duration is T_D, and the offset edge is the reverse of the onset edge making the window symmetrical.

As for the rectangular window, the Fourier transform of the raised-cosine window can be written in terms of a real function multiplied by a time-delay phase factor. A half dozen pages of algebra suffice to show that the Fourier transform is

$$W_{\mathrm{Hann}}(\omega) = \mathrm{Hann}(\omega)\, e^{-i\omega T_D/2}, \quad (16)$$

where Hann is the transform of the symmetrical window, again involving the sinc function,

$$\mathrm{Hann}(\omega) = \frac{2\cos(\omega\tau_o/2)}{1-(\omega\tau_o/\pi)^2} \cdot \frac{\sin[\omega(T_D-\tau_o)/2]}{\omega}. \quad (17)$$

Function $\mathrm{Hann}(2\pi f)/(T_D - \tau_o)$ is plotted in Fig. 2(b).

For windows other than rectangular, such as the raised-cosine window, the significant duration is not equal to the total duration T_D. For the raised-cosine window, Eq. (17) shows that the significant duration is $T_S = T_D - \tau_o$. Therefore, the fundamental angular frequency of the unlimited-duration waveform needs to be $\omega_o = 2\pi/(T_D - \tau_o)$ in order for the window and the waveform to be matched. That result was not immediately obvious.

C. Trapezoid window

The trapezoid window is shown in Fig. 3(a). It has straight-line onset and offset edges, both with duration τ_o. The Fourier transform of the trapezoid (less than half a dozen pages of algebra) is given by Eq. (18) and is shown in Fig. 3(b).

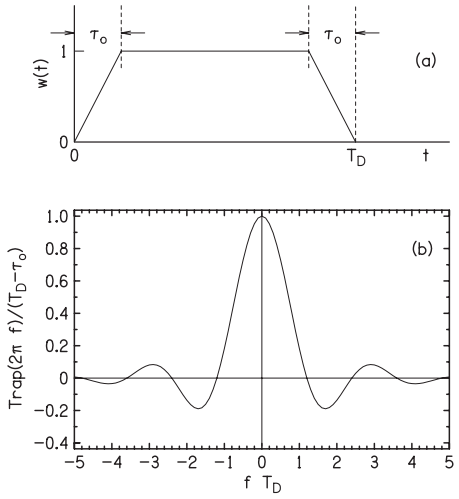


FIG. 3. Trapezoid temporal window for $\tau_o = T_D/6$. (a) The window as a function of time. The full-on duration is $T_D - 2\tau_o$. (b) The Fourier transform of the trapezoid window apart from the phase factor. The Fourier transform is zero when fT_D is somewhat larger than integer values.

$$W_{\mathrm{Trap}}(\omega) = \mathrm{Trap}(\omega)\, e^{-i\omega T_D/2}, \tag{18}$$

where "Trap" is the transform of the symmetrical trapezoid window,

$$\mathrm{Trap}(\omega) = \frac{4\,\sin(\omega\tau_o/2)}{\omega\tau_o} \cdot \frac{\sin[\omega(T_D - \tau_o)/2]}{\omega}. \tag{19}$$

It is evident that the significant duration for the trapezoid window is again $T_S = T_D - \tau_o$.

The different temporal windows described above lead to different amounts of spectral splatter. A good predictor for splatter is the high-frequency asymptotic behavior of the Fourier transforms. As shown in Eqs. (11), (19), and (17), respectively, the rectangular window spectrum decreases as ω^{-1}, the trapezoid window spectrum decreases as ω^{-2}, and the raised-cosine window spectrum decreases as ω^{-3}. These asymptotic behaviors cannot be seen well on the spectral plots in Figs. 1–3 because the horizontal axis does not extend to high frequencies.

III. MATCHING AND MISMATCHING

A. Alternative representations

Sections I and II can be interpreted as follows: When a signal is created by windowing an unlimited-duration waveform, the spectrum becomes a continuous function of frequency. However, there are several possible discrete-frequency representations of this spectrum. One representation begins with a long analysis interval, T_L, longer, perhaps much longer, than T_D. This long time interval is padded with zeros outside the signal interval. The spectrum will consist of lines at frequencies n/T_L. It will include the window information and will also force the signal to be zero outside the signal interval. This representation contains the most information, and it incorporates the fact that the signal does not live forever. However, it does not resemble the spectrum of the waveform. In this representation, much of the spectral power for major components of the waveform can reside in frequencies very different from the frequencies in the waveform.

A second representation periodically extends the signal as windowed. If the window has overall duration T_D, the Fourier transform of the periodically-extended signal exists only on harmonics of a fundamental frequency $2\pi/T_D$. This Fourier transform includes information about the window shape as an integral part of the signal. To the extent that window $W(\omega)$ decreases with increasing ω, this Fourier representation is band limited. However, the spectrum of the periodically-extended signal does not agree with the spectrum of the original unlimited-duration waveform because the window is not matched to the waveform. Also, because the periodically-extended signal is an artifice, which enables the discrete-frequency representation but does not force the waveform to zero outside the signal interval, this Fourier representation is incomplete.

A third representation of the signal is the matching approach suggested in this article. The Fourier transform of the window indicates the significant duration T_S. If the unlimited-duration waveform consists of harmonics of fundamental frequency $\omega_o = 2\pi/T_S$, then the waveform is matched to the window. Then the spectrum of the signal evaluated on these harmonics is the same as the spectrum of the unlimited-duration waveform. Matching the window and the waveform causes the window to disappear in this representation, and the spectrum is said to be under control. As an example of this control, the spectrum of a *filtered windowed* waveform becomes the same as for a *windowed filtered* waveform, as shown in the Appendix. The order of windowing and filtering operations does not matter if window and waveform are matched.

The rectangular window is a special case. When a rectangularly-windowed signal is periodically extended, it becomes equivalent to the waveform that matches the window. For the rectangular window, one can have both a continuous periodic extension and waveform matching. However, the continuous spectrum may not be sufficiently band limited to avoid distracting transients.

B. An example with three steps

Spectral effects of temporal windows that are matched or mismatched to the waveform can be illustrated by a simple example with five spectral components. We suppose that the unlimited-duration waveform is a narrow noise band, 40 Hz wide centered at 500 Hz, with amplitudes and phases (expressed in degrees) chosen haphazardly:

$$\begin{aligned} x(t) = {} & 0.3\cos(360\cdot 480t - 45) + 1.0\cos(360\cdot 490t - 5) \\ & + 0.2\cos(360\cdot 500t + 17) + 0.5\cos(360\cdot 510t + 143) \\ & + 0.8\cos(360\cdot 520t - 17). \end{aligned} \tag{20}$$

The spectrum of the unlimited-duration waveform is shown in Table I, row (a).

The fundamental frequency of the unlimited-duration waveform in Eq. (20) is 10 Hz because we want to make a noise that is approximately 100 ms long. In a first step, we

TABLE I. Spectra for matched and mismatched windowed signals are given in the form level (dB) | phase (degrees). (a) Spectrum of the standard unlimited-duration signal from Eq. (20). The experimenter wants to preserve this spectrum. The amplitudes from the second line are converted to levels in decibels for better comparison with other parts of the table. Step 1: Row (a) is also the spectrum of the rectangularly-windowed signal. Step 2: Row (b) is the spectrum of the windowed signal with mismatched window and waveform. Step 3: Row (c) is the spectrum of the windowed signal with matched window and waveform. Row (d) is the spectrum of the windowed signal with the 500-Hz component shifted by 180° using mismatched window and waveform. The levels are computed with respect to the largest component in row (b) to make the binaural comparison correct. Row (e) is the spectrum of the windowed signal with the 500-Hz component shifted by 180° using a matched window and waveform.

Frequency (Hz)	480	490	500	510	520
Amplitude	0.3	1.0	0.2	0.5	0.8
(a) Standard:	−10.5 | −45	0.0 | −5	−14.0 | 17	−6.0 | 143	−1.9 | −17
(b) Mismatched:	−12.2 | −73	0.0 | −5	−20.7 | 70	−2.3 | 151	−2.1 | −20
(c) Matched:	−10.5 | −45	0.0 | −5	−14.0 | 17	−6.0 | 143	−1.9 | −17
(d) 180 - Mismatched:	−12.1 | −63	+0.4 | −4	−8.3 | −174	−2.7 | 148	−2.1 | −18
(e) 180 - Matched:	−10.5 | −45	0.0 | −5	−14.0 | −163	−6.0 | 143	−1.9 | −17

use a rectangular window with a duration $T_D=100$ ms. Because the window and the waveform are matched, the spectrum of the 100-ms noise preserves the amplitudes and phases of the unlimited-duration waveform. The Fourier series spectrum, equivalent to the spectrum of the periodically-extended signal, is equal to the spectrum of the unlimited-duration waveform and is again given by Table I, row (a).

In a second step, we apply a 10-ms raised-cosine edge to the beginning and end of the 100-ms noise, as shown in Fig. 4(a). The total duration remains $T_D=100$ ms, the full-on duration becomes 80 ms, and the significant duration in Eq. (17) becomes $T_D-\tau_o=90$ ms. This value of significant duration would match a waveform having a fundamental frequency of 1000/90 Hz, but it does not match our chosen fundamental frequency of 10 Hz. The Fourier series spectrum for harmonics of 10 Hz is shown in Table I, row (b). It does not look good. Both the amplitude spectrum and the phase spectrum are distorted because these spectra are trying to capture some elements of the window. In addition, there is spectral splatter for harmonics of 10 Hz outside the original 40-Hz band not shown in the table. Further, the spectrum $Y(\omega)$ looks no better on a different set of frequencies because the waveform, with its fundamental of 10 Hz, is fundamentally incompatible with this raised-cosine window. By mismatching window and waveform, we have lost precise control of the final spectrum.

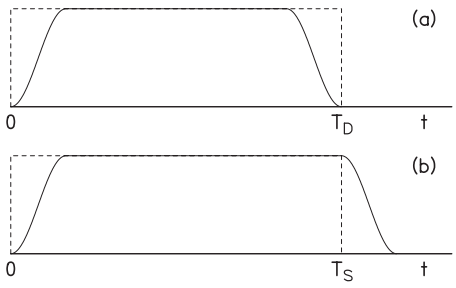


FIG. 4. Mismatched and matched windows. (a) The rectangular window (dashed) matches a waveform with fundamental frequency $1/T_D$. The raised-cosine window does not match. (b) Both the rectangular window and the raised-cosine window match a waveform with fundamental frequency $1/T_S$, where T_S is the significant duration.

In a third step, we retain the 10-ms raised-cosine edge at the beginning and end, but we let the total duration be $T_D=110$ ms, as shown in Fig. 4(b). Then the significant duration is 100 ms, and that matches a fundamental frequency of 10 Hz. The amplitude and phase spectra, computed at multiples of 10 Hz, are shown in Table I, row (c). Table I, row (c) is identical to Table I, row (a), and there is no spectral splatter for harmonics of 10 Hz outside the original 40-Hz band. What we have given up in exchange for a controlled spectrum is the periodically-extended signal. The 10-Hz fundamental frequency is incompatible with periodic extension of the 110-ms signal.

C. Binaural consequences

Because research in binaural hearing requires signals with precise interaural amplitude and phase properties, it is interesting to investigate the binaural consequences of windowing waveforms. In this investigation, we imagine that we intend to synthesize signals for the left and right ears where the interaural differences are specified in the spectra of the unlimited-duration waveforms. We would like to know about the effects on these interaural differences if we apply a temporal window to the waveforms. The following facts apply.

- If the period of the waveforms is equal to the significant duration of the temporal window, or an integral submultiple of it, then the amplitudes and phases on waveform frequencies in both the left and right channels are unchanged by windowing. Consequently, the interaural properties, interaural phase difference (IPD), and interaural level difference (ILD), after the window is applied, are the same as the interaural properties of the waveforms on those frequencies. Life is good binaurally.
- If the only interaural differences are an IPD ($\Delta\phi$) and an ILD (g) applied identically to all components in the spectrum of the unlimited-duration signal, then the interaural spectral properties of the signal after the window is applied are the same as for the unlimited-duration signal whatever the temporal window. It takes only a few lines of mathematics to prove this fact, starting with Eq. (8). If the waveform

in the left ear is $x_L(t)$, then the windowed-signal spectrum in the left ear is

$$Y_L(\omega) = \frac{1}{2\pi} \int d\omega' W(\omega - \omega') X_L(\omega'), \tag{21}$$

and the windowed-signal spectrum in the right ear is

$$Y_R(\omega) = \frac{1}{2\pi} \int d\omega' W(\omega - \omega') [g(\omega') e^{i\Delta\phi(\omega')}] X_L(\omega'), \tag{22}$$

where g is the gain leading to the ILD and $\Delta\phi$ is the interaural phase shift.

By hypothesis, the gain and interaural phase shift are independent of frequency ω', and they can be extracted from the integral. Consequently,

$$Y_R(\omega) = (g e^{i\Delta\phi}) Y_L(\omega), \tag{23}$$

which says that the signals after windowing have the same interaural relationships as the unlimited-duration waveforms. This binaural invariance always holds good, whether or not the waveform and window are matched. Again, life is good binaurally.

• If the ILD or the IPD is not the same for all frequencies, then a mismatch between the waveform and the window leads to a distorted interaural spectrum. The effects may or may not be important. For instance, if an interaural time difference (ITD) is applied, the IPD is different for different frequencies, but it may not be very different, especially for a narrow-band noise. Given an applied ITD in the range of a typical experiment (<1 ms) and a critical-band noise, the spectral distortion may not be severe. It depends partly on the number of components in the waveform and partly on luck.

By contrast, a dramatic phase change can lead to dramatic distortion. For instance, one might try to create an NoSπ stimulus by synthesizing two channels that are identical except that the phase of one component is reversed by 180°. Then a mismatch between waveform and window [Fig. 4(a)] leads to serious distortion, where the interaural amplitudes and phases of the windowed signal do not resemble the desired stimulus.

For example, beginning with the five-component waveform of Eq. (20), and reversing the phase of the center component (500 Hz) leads to the interaural differences shown in Table I, row (d). The IPDs for the five components are expected to be 0°, 0°, 180°, 0°, and 0°. The actual IPDs for the mismatched condition can be obtained by subtracting Table I, row (b) from Table I, row (d). They are 10°, 1°, 116°, −3°, and 2°. The expected phase shift of 180° has been turned into 116°. Further, reversing the phase of the central component has changed the amplitudes of the components. The level of the central component has changed by more than 12 dB. That was not at all intended.

By contrast, if the matched window shown in Fig. 4(b) is used then reversing the phase of the central component leads to the spectrum shown in Table I, row (e). Comparison with Table I, row (c) shows that the only interaural difference is the 180° phase shift of the central component as expected. To create a controllable stimulus like this, it is essential that the waveform and window be matched.

D. Size of the distortion

The discrepancies between the spectrum of the mismatched windowed signal and the spectrum of the unlimited-duration waveform on the set of waveform frequencies were examined for the particular waveform given in Eq. (20). That waveform had a relatively small amplitude for the central component, and it is partly for this reason that the desired IPD of 180° was so badly violated in the windowed binaural signal. It became 116°. That kind of result is expected. When the waveform and window are not matched, the amplitude of any particular component in the windowed signal becomes a linear combination of the amplitudes of all the other components, more or less, depending on all the phases. A component of the original unlimited-duration signal that is small is particularly vulnerable to distortion by other components with larger amplitudes. Although the distortion seen for this particular waveform is large, it is not atypical for this case of mismatched waveform and window. Other choices of original amplitudes and phases lead to distortions that are as large or larger.

In the example above, the waveform and window are only somewhat mismatched. Because the window has a total duration $T_D = 100$ ms and a raised-cosine edge of $\tau_o = 10$ ms, the edges of the window account for only 20% of the total window. The window is not greatly different from a rectangular window, which would preserve the spectrum.

If the calculations leading to Table I are repeated except that the edge duration is increased to $\tau_o = 20$ ms, then the spectral distortion is correspondingly more dramatic. For instance, in the windowed *binaural* signal the IPD of the central component becomes −43° degrees instead of 180°. Generally, as the edge duration becomes a greater fraction of the total duration, the effect of mismatch becomes larger.

Because the distortion of the spectrum for a given component in a mismatched case is a linear combination of amplitudes for all the other components, one expects the distortion to be larger when there are more components. Thus, the distortion observed in the above example with only five components may underestimate the typical distortion. However, the form of $W(\omega)$ shows that the coefficients in the linear combination decrease as the component contributing to distortion is farther away from the particular component of interest. For the raised-cosine window, it decreases as the cube of the distance in frequency, which is better than the trapezoid window.

IV. PRACTICAL CONSIDERATIONS

A. Tones

Section III above, including the five-component example, dealt with a dense spectrum, where the spacing between adjacent components in the signal was as small as was allowed by the window, namely, $2\pi/T_S$. It is, of course, possible to use only a subset of the allowed set of frequencies to create a tone. A complex tone with fundamental angular fre-

quency $M(2\pi/T_S)$, second harmonic $2M(2\pi/T_S)$, and so on with integer M is also matched to the window with significant duration T_S.

B. Discrete Fourier transform

Thus far, the matching principle has been developed in terms of signals that are continuous functions of time. The principle also applies to discrete time (sampled) signals as well and to the discrete Fourier transform (DFT). As noted in the Introduction, an experimenter normally creates a signal beginning with an intended spectrum. In a digital implementation, the experimenter chooses a sample rate and a signal duration so that the intended spectrum can be represented as a spectral array (Proakis and Manolakis, 1992). Applying an inverse DFT, or inverse fast Fourier transform, produces a function of time, which is then given a temporal window to make the signal. This approach to signal generation does not match the waveform and the window.

In order to match the waveform and the window, the spectral array should be represented assuming that the signal duration will be the significant duration T_S, not the signal duration T_D. As a result, the inverse DFT will lead to a function that has duration T_S. That duration is not long enough because the duration of the intended signal, T_D, is always longer than T_S. For instance, for the raised-cosine window, the duration is $T_D=T_S+\tau_o$. However, by its nature, the inverse DFT produces a function of time that is periodically extended. Consequently, it is only necessary to repeat a portion of that function in order to produce a function with duration T_D. That function can then be multiplied by the temporal window, and the final signal will correspond to a matched waveform and window.

V. DISCUSSION

Experimenters initially specify their signals in terms of a desired spectrum. Precise spectral requirements in terms of discrete frequencies imply a waveform of unlimited duration. For instance, an experimenter might specify a 500-Hz sine tone or a noise with a rectangular bandwidth of 40 Hz. These are statements about infinitely long waveforms, but a signal as actually used in real life has a finite duration, imposed by a temporal window. Making the duration finite inevitably has spectral consequences. The windowed signal, as presented to a listener, does not have the spectrum as specified.

Summary of matching. This article has considered temporal windowing procedures that are spectrum preserving and windowing procedures that are not spectrum preserving. The spectrum to be preserved (or not) is that of a periodic waveform, where the spectrum exists only on a set of harmonically related frequencies $\{f_n=nf_o\}$. The rectangular window with duration T_D preserves the spectrum if $f_oT_D=1$. On the set of harmonic frequencies $\{nf_o\}$, the rectangularly-windowed signal has no power outside the band of desired spectral components; i.e., there is no spectral splatter on harmonics. Further, within the band of desired components the spectrum of the windowed signal has exactly the amplitudes and phases of a periodically-extended signal. The rectangular window is conceptually simple because the matched unlimited-duration waveform is the same as the periodically-extended signal.

The purpose of this article was to point out that spectral preservation arises from properties of the Fourier transform of the temporal window. It was shown that if the window and waveform are matched, then the spectrum is preserved. The rectangular window with duration T_D is matched to the waveform made from spectral components with frequencies that are harmonics of $1/T_D$. If some other form of temporal window, having overall duration of T_D, is used, then the window is not matched by a waveform made from harmonics of $1/T_D$. Instead, the match between window and waveform must be made in view of the significant duration T_S that appears in the Fourier transform of the window. For example, the raised-cosine window shown in Fig. 2 having a total duration of T_D and a Hanning edge with duration τ_o at each end is matched by a waveform with period $T_S=T_D-\tau_o$. A waveform that is matched to a window that is not rectangular is not equal to the periodically-extended signal with period T_D. Given that the periodically-extended signal is only a fiction, extrapolated from the Fourier series, giving it up entails little cost.

Value of matching. Because matching the waveform and the window only preserves the spectrum on a specific set of frequencies, one may well ask whether there is a value to matching. One value is that the frequencies of this set are the most important frequencies in the windowed signal. They are almost certain to be the frequencies with the largest amplitudes and the most power. Also, they are the frequencies that are initially specified by the experimenter.

However, sometimes experimenters make no attempt to match. For instance, an experimenter may compute a noise waveform using a band of spectral components spaced by only 1 Hz, give it a duration (e.g., 100 ms) using a nonrectangular window of some form, and present it to a listener. If the window is smooth, the spectral splatter outside the band may be held to some designated limits. Although the phases and amplitudes within the band of the windowed noise are not preserved, nor are they equal to those of the periodically-extended waveform with a period of 1 s, that may not be of much concern to the experimenter. It may be that one set of amplitudes and phases is as good as any other.

Creating waveforms in this way, with no regard for matching, leads to the widest possible variety of signals in the experiment. However, the spectrum is not well controlled. Generating noise stimuli in this way is like using a thermal noise generator and bandpass filter for the waveform and an analog multiplier for the window. Stimuli like these would not be appropriate for reproducible noise experiments where the stimulus spectrum needs to be known exactly. An alternative procedure for reproducible noise is to begin with a desired spectrum, create a windowed stimulus violating all the matching rules, and then test the windowed stimuli for approximate agreement with the desired spectral properties. That approach has been taken by a number of experimenters, e.g., Goupell and Hartmann, 2007. It is normally necessary to reject a large number of candidate stimuli.

Generalization. The mathematical development in this article considered three temporal windows, rectangular,

raised cosine, and trapezoidal. Conditions for matching to a waveform were found for all three of those windows. Is it possible to generalize this development? Can one say that for every form of temporal window there is some way to obtain a matching condition? Apparently not. The matching conditions for the three windows depended on Fourier transforms that contained a sinc-function factor. On a set of matching harmonics, the sinc function became a Kronecker delta. Other windows do not have a sinc-function factor, a Gaussian window, for example. The matching conditions as derived in this article would not work for such a window.

Signal detection theory. In the theory of signal detection, the number of degrees of freedom in a signal is given by $2W_bT_D$, where W_b is the bandwidth (Green and Swets, 1966). This result implicitly assumes a rectangular temporal window. That can be seen as follows: For a rectangular window, the frequency spacing in the spectrum of the periodically-extended signal is $1/T_D$. The number of independent spectral components is then simply the bandwidth divided by the frequency spacing. The factor of 2 arises because specifying each component requires two variables, an amplitude and a phase. According to the theory of this article, if the window is not rectangular, the number of degrees of freedom is smaller. For instance, for a raised-cosine window, the number of degrees of freedom is $2W_b(T_D-\tau_o)$.

The ongoing signal. This article has been concerned with the Fourier transform of a windowed signal. The signal originates in a waveform of unlimited duration, but in the end, the window, of whatever form, is considered to be an integral part of the signal being Fourier transformed. That point of view is not necessarily perceptually relevant. It may be that the listener's decisions are affected only by the ongoing portion of the signal. For instance, if the listener's task is to evaluate the pitch of a one-second periodic complex tone, it is unlikely to matter whether the stimulus is turned on with one kind of temporal window or another. Then the most relevant spectral representation would be the spectrum of the unlimited-duration waveform. Spectral distortions caused by a mismatched raised-cosine window would be unimportant because that window plays a role that is merely cosmetic. Its smooth edges eliminate unaesthetic clicks.

Relevance of the spectrum. For some stimuli, particularly those that are very short, the temporal window contributes importantly to the power spectrum of the signal presented to a listener. Sometimes considering the entire power spectrum leads to a insights into perception (e.g., Green, 1968; Hartmann and Sartor, 1991). However, the power spectrum may not always provide the most relevant physiological or psychological representation of the stimulus. That is particularly true for the simple spectrum obtained by matching the waveform and window as described in this article. What is one to make of the fact that on one set of frequencies there is spectral splatter but on another set of frequencies there is none? Alternatives to the spectral representation are available, e.g., the wavelet or Wigner distribution, but they have had negligible impact on auditory science compared to the Fourier spectral representation, which maps to place in the auditory system.

Another alternative to a spectral representation is to build a mathematical model of the system of interest—anything from outer ear to cortex—and to use the time-dependent windowed signal $y(t)$ as the input to the model. Whether the essential stimulus attributes lie in the onset, offset, window, or ongoing signal then becomes a characteristic of the model. The spectrum itself plays no role.

Final word. When it is important to control the spectrum of a windowed stimulus, there is a value to matching the waveform and the window. To match the window and waveform it is necessary to know the Fourier transform of the windowing function. Control may be particularly important in some binaural experiments where differences in the signals to the two ears become important.

ACKNOWLEDGMENTS

We are grateful to Dr. H. S. Colburn for comments on this manuscript. The manuscript was written while W.M.H. was a visitor in the Department of Biomedical Engineering at Boston University. E.M.W. was supported by the Michigan State University Undergraduate Professorial Assistant Program. This work was supported by the NIDCD Grant No. DC-00181.

APPENDIX: FILTERED FUNCTIONS

If waveform and window are matched, the spectrum of a windowed filtered signal is the same as the spectrum of a filtered windowed signal, where the spectrum is defined on the discrete set of frequencies that are harmonics in the waveform. The order of operations does not matter. This appendix proves that fact. It begins with the windowed filtered signal.

If the waveform $x(t)$ is filtered with transfer function $H(\omega)$, the Fourier transform of the filtered waveform is $H(\omega)X(\omega)$. If the filtered waveform is then windowed with temporal window $w(t)$, the Fourier transform is given by the convolution from Eq. (8),

$$Y_1(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}d\omega' W(\omega-\omega')H(\omega')X(\omega'). \tag{A1}$$

Because Eq. (3) expresses $X(\omega)$ as a sum of delta functions, the integral is easy to do, and

$$Y_1(\omega)=\frac{1}{2}\sum_{n=-\infty}^{\infty} W(\omega-\omega_n)H(\omega_n)X_n. \tag{A2}$$

If $x(t)$ is periodic with fundamental frequency ω_o, and if Y_1 is evaluated on harmonics of ω_o, then

$$Y_1(m\omega_o)=\frac{1}{2}\sum_{n=-\infty}^{\infty} W[(m-n)\omega_o]H(n\omega_o)X_n. \tag{A3}$$

Reversing the order of the operations above produces a filtered windowed signal. Beginning with Eq. (8) for the windowed waveform and then filtering with transfer function H

leads to the Fourier transformed signal

$$Y_2(\omega) = \frac{1}{2\pi} H(\omega) \int_{-\infty}^{\infty} d\omega' W(\omega - \omega') X(\omega'). \quad \text{(A4)}$$

Using Eq. (3) to write $X(\omega')$ in terms of discrete harmonic frequencies and again evaluating at those frequencies lead to

$$Y_2(m\omega_o) = \frac{1}{2} H(m\omega_o) \sum_{n=-\infty}^{\infty} W[(m-n)\omega_o] X_n. \quad \text{(A5)}$$

According to the development in this appendix, ω_o is the fundamental frequency of the waveform. If the window is matched to the waveform, then $W[(m-n)\omega_o] = W_o \delta_{m,n}$, where W_o is a constant. Then Eq. (A3) for Y_1 and Eq. (A5) for Y_2 are the same, and

$$Y_2(m\omega_o) = Y_1(m\omega_o) = \tfrac{1}{2} H(m\omega_o) W_o X_m. \quad \text{(A6)}$$

The filtered windowed signal is equal to the windowed filtered signal. If the waveform and the window are not matched, $W[(m-n)\omega_o]$ is not diagonal on m and n and the filtered windowed signal is different from the windowed filtered signal.

Goupell, M. J., and Hartmann, W. M. (**2007**). "Interaural fluctuations and the detection of interaural incoherence II: Brief duration noises," J. Acoust. Soc. Am. **121**, 2127–2136.

Green, D. M. (**1968**). "Sine and cosine masking," J. Acoust. Soc. Am. **44**, 168–175.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Hartmann, W. M. (**1998**). *Signals, Sound, and Sensation* (Springer, New York).

Hartmann, W. M., and Sartor, D. (**1991**). "Turning on a tone," J. Acoust. Soc. Am. **90**, 866–873.

Proakis, J. G., and Manolakis, D. G. (**1992**). *Digital Signal Processing: Principles, Algorithms, and Applications*, 2nd ed. (Macmillan, New York).

Automatic detection of articulation disorders in children with cleft lip and palate

Andreas Maier, Florian Hönig, Tobias Bocklet, and Elmar Nöth[a)]
Lehrstuhl für Informatik 5 (Mustererkennung), Universität Erlangen-Nürnberg, 91058 Erlangen, Germany

Florian Stelzle and Emeka Nkenke
Mund- Kiefer- und Gesichtschirurgische Klinik, Universitätsklinikum Erlangen, 91054 Erlangen, Germany

Maria Schuster
Abteilung für Phoniatrie und Pädaudiologie, Universitätsklinikum Erlangen, 91054 Erlangen, Germany

(Received 5 November 2008; revised 1 August 2009; accepted 6 August 2009)

Speech of children with cleft lip and palate (CLP) is sometimes still disordered even after adequate surgical and nonsurgical therapies. Such speech shows complex articulation disorders, which are usually assessed perceptually, consuming time and manpower. Hence, there is a need for an easy to apply and reliable automatic method. To create a reference for an automatic system, speech data of 58 children with CLP were assessed perceptually by experienced speech therapists for characteristic phonetic disorders at the phoneme level. The first part of the article aims to detect such characteristics by a semiautomatic procedure and the second to evaluate a fully automatic, thus simple, procedure. The methods are based on a combination of speech processing algorithms. The semiautomatic method achieves moderate to good agreement ($\kappa \approx 0.6$) for the detection of all phonetic disorders. On a speaker level, significant correlations between the perceptual evaluation and the automatic system of 0.89 are obtained. The fully automatic system yields a correlation on the speaker level of 0.81 to the perceptual evaluation. This correlation is in the range of the inter-rater correlation of the listeners. The automatic speech evaluation is able to detect phonetic disorders at an experts'level without any additional human postprocessing.
 [DOI: 10.1121/1.3216913]

PACS number(s): 43.70.Dn, 43.72.Ar, 43.80.Qf, 43.80.Vj [DOS] Pages: 2589–2602

I. INTRODUCTION

Communication disorders are a major challenge in the 21st century because of their personal and financial impact. The cost of care as well as the decrease in employment opportunities for people with communication disorders cause a loss of $154 billion to $186 billion per year to the economy of the United States of America alone.[1] People with speech disorders do not only suffer from restricted speech but also from vocational limitations. The use of automatic speech processing techniques can contribute to reduce the associated costs. More specifically, such methods can affect speech screening and therapy as follows.

- Speech processing can serve as an easy-to-apply diagnostic tool and can also be used for speech screening. The cost of diagnosis can be reduced with such an automatic system because it can also be performed by nonprofessionals.
- Therapy strategies can be evaluated and compared against each other in clinical trials or for individual therapy.
- Speech processing can support therapy sessions in the practice as well as telemedical therapy sessions, which can be performed by the patient from his home.

In this work we focus on speech attributes related to cleft lip and palate (CLP). CLP might cause communication disorders, especially articulation disorders. CLP is the most common malformation of the head. It constitutes almost two-thirds of the major facial defects and almost 80% of all orofacial clefts.[2] Its prevalence differs in different populations. CLP appears most often in Asians with a prevalence of 1 in 400–500 newborns and least often in African Americans with 1 in 1500–2000 newborns.[3,4] Speech of children with CLP is sometimes still disordered even after surgery and might show special characteristics such as hypernasality (HN), backing, and weakening of consonants.[5]

The major feature of disordered speech in CLP is HN in vowels (perceived as characteristic "nasality") and nasalized consonants (NC). This may reduce the speech intelligibility.[6–8] Both features, HN and NC, can be summarized as nasal air emission.

The term nasality is often used in the literature for two different kinds of nasality: HN and hyponasality. While HN is caused by enhanced nasal emissions, as in CLP children, hyponasality is caused by a blockage of the nasal airway, e.g., when a patient has a cold. There are several studies on both nasality types.[9] However, most of them concern only the effects on voiced speech (vowels)[10–12] and consonant-vowel combinations.[13,14]

Figure 1 shows the effect of nasalization in the envelope spectrum[15] of vowel /a:/. In both spectra a slight nasal formant $F_1^N(f)$ exists between at frequency $f=300$ and 500 Hz.

[a)]Author to whom correspondence should be addressed. Electronic mail: andreas.maier@cs.fau.de

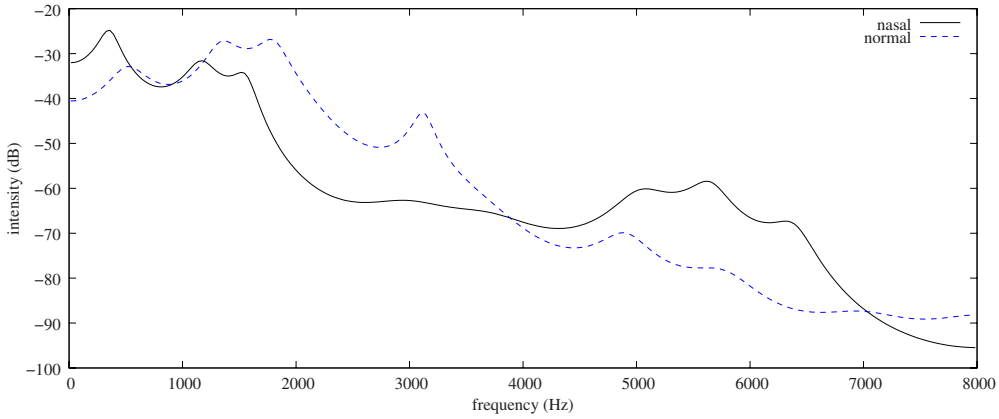


FIG. 1. (Color online) LP-model spectrum of a nasal and a non-nasal realization of the phoneme /a:/ in the phonetic context /ha:s@/ ("Hase," the German word for "hare") using 20 LPC-coefficients: The intensity of the nasal formant $F_1^N(f)$ ($f \approx 300-500$ Hz) is stronger than the intensity of the first formant $F_1(f)$ ($f \approx 1100-1300$ Hz) in the nasal realization. Note that the displayed speech is children's speech, which causes exceptionally high formant frequencies.

The maximal intensity of the first formant $F_1(f)$ is at about 1100–1300 Hz. In the nasalized /a:/, the intensity of the $F_1^N(f)$ is stronger than the $F_1(f)$, which makes the nasality audible. Actually, this effect is caused by a combination of the following effects.[16]

- The first formant bandwidth increases while the intensity decreases.
- The nasal formant $F_1^N(f)$ emerges or is increased.
- Antiresonances appear, which increase the strength of the so-called antiformants $F_k^A(f)$.

According to the literature, the main cause for audible nasality is the intensity reduction in the first format.[16,17]

Nasality in consonants, however, shows different acoustic properties depending on their mode of articulation, e.g., voiced or unvoiced. Effects in the formant structure can only be analyzed in the neighboring vowels. The effects on the consonants, however, are still audible. In fricatives, for example, the nasality is audible as a general weakening of the energy of the phoneme with additional streaming noises caused by the nasal air flow. In contrast to the non-nasal consonant the way to the nasal cavity is open. Hence, at least some of the emitted air flows through the nose and the amount of air that is emitted through the mouth is reduced. In the literature these effects are rarely described and often only the analysis of vowel-consonant clusters is performed.[18]

The speech of CLP children might also contain secondary cleft-type characteristics. These originate from compensatory articulation, which may still be present even after adequate closure of the clefting. For example, pharyngeal backing (PB) is caused by a shift in the localization of the tongue toward the palate during the articulation. Glottal articulation [also called laryngeal replacement (LR)] is an extreme backing of articulation. The resulting acoustic realization is similar to that of a glottal stop. Another typical characteristic of secondary phonetic disorders is the absence or weakening of consonants[5] [weakened plosives (WPs)].

In clinical practice, articulation disorders are mainly evaluated perceptually, and the evaluation procedures are mostly performed by a speech therapist. Previous studies have shown that experience is an important factor that influences the judgment of speech disorders. The perceptual evaluation of persons with limited experience tends to vary considerably.[19,20] For scientific purposes, usually the mean score judged by a panel of experienced speech therapists serves as a reliable evaluation of speech and is sometimes called "objective." Of course, this is very time and manpower consuming. Until now, objective measures only exist for nasal emissions[7,9] and for voice disorders in isolated vowels.[17,21] But other specific articulation disorders in CLP cannot be reliably and objectively quantified yet. In this paper, we present a new technical procedure for the objective measurement and evaluation of phonetic disorders in connected speech, and we compare the obtained results with perceptual ratings of an experienced speech therapist. We present two experiments.

- In a first experiment an automatic speech recognition (ASR) system was applied to evaluate the detection of the above mentioned articulatory features of CLP speech (HN, NC, PB, LR, and WP). The experiment is based on the transliteration of the tests that was created manually.
- A second experiment was conducted to examine whether it is possible to perform the assessment fully automatically without manual transliteration.

II. SPEECH DATA

58 children with CLP were recorded during the commonly used PLAKSS speech test (psycholinguistische analyse kindlicher sprechstörungen — psycholinguistic analysis of children's speech disorders). The acoustic speech signal was sampled at 16 kHz with a quantization of 16 bits. Informed consent had been given by the parents prior to all recordings.

For the first experiment recordings of 26 children at an age of 9.4 ± 3.3 years were used (CLP-1). Two of the children in the data set had an isolated cleft lip, 3 an isolated cleft palate, 19 unilateral CLP, and 2 bilateral CLP. The recordings were made with a head set (dnt Call 4U Comfort) and a standard PC.

The recordings were performed in the same manner as during the therapy session: The test was presented on paper-

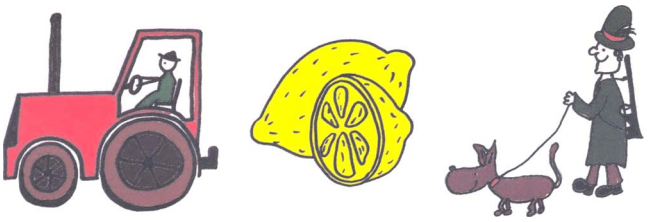


FIG. 2. (Color online) Slide 13 of the PLAKSS test: "Trecker, Zitrone, Jäger" (tractor, lemon, hunter).

board and stored in a single audio file. Therefore, the time stamps at which the therapist switched from one slide to another was not known. The data set is a subset of the data that were investigated in a previous study concerning speech intelligibility[22] and semiautomatic evaluation of speech disorders.[23]

The second group (CLP-2) was formed by 32 children at the age of 8.7 ± 1.7 years. Five of the children had a cleft lip, 7 a cleft palate, and 20 a unilateral CLP. No child in the data set had a bilateral cleft. They were recorded directly at the PC. The PC was used to display the slides and to perform the recording simultaneously. The audio data of each slide are stored in an individual audio file. Hence, the correspondence of audio data and the respective slide is clear. Furthermore, we presented the correct target words in small letters at the bottom of the screen in order to decrease the variability in the test data. The data were recorded and evaluated as described in the following paragraph using the program for the evaluation of all kinds of speech[24] (PEAKS) disorders.

The PLAKSS test[25] is a German semistandardized test for articulation disorders. It consists of 99 pictograms (97 disjoint) which are shown on 33 slides. It was designed to record also young children who are not yet able to read. The test contains all German phonemes in different positions (word initial, central, and final).

Figure 2 shows an example of the slides. It depicts the German words "Trecker, Zitrone, Jäger" to test for the phoneme /r/ in consonant-consonant clusters and at the end of a word. The words mean tractor, lemon, and hunter in English. It gives a good example: While the tractor and the lemon are quite easy to identify, the hunter often poses a problem. Many children do not recognize the rifle on the back of the hunter and call the pictogram "man with a dog." Furthermore, the word "Trecker" is rather uncommon in the southern part of Germany. Children tend to prefer variants such as "Traktor" or "Bulldog." Therefore, the vocabulary of the PLAKSS test has to be extended with common word alternatives and regional variants if their automatic detection is desired.

As the test has to be performed by a supervisor who gives instructions during the test, the voice of the supervisor is always also audible on the audio tracks.

III. SEMI- AND FULLY AUTOMATIC SEGMENTATION

In both data sets CLP-1 and CLP-2 the data were segmented using an ASR system. We use an ASR system based on hidden Markov models (HMMs). It is a word recognition (WR) system developed at the Chair of Pattern Recognition (Lehrstuhl für Mustererkennung) of the University of Erlangen-Nuremberg. In this study, the latest version[24] was used.

As the performance of speech recognition is known to be dependent on age,[26] several recognizers were trained for certain age groups. According to previous evaluations,[27] the best groups for the creation of age-dependent recognizers were found to be

- <7 years,
- 7 years,
- 8 years,
- $9+10$ years, and
- >10 years.

A maximum likelihood linear regression (MLLR) adaptation was performed on the acoustic models using the HMM output probabilities[28–30] in order to improve the recognition for each child.

The CLP-1 data set was segmented semiautomatically using the transliteration of the speech data. In the CLP-2 database this step was replaced by a fully automatic procedure using PEAKS.[24] These segmentation procedures are described in the following.

A. Semiautomatic segmentation procedure

As the whole speech data of one child were collected in a single audio file in the CLP-1 data, the complete data set had to be transliterated in order to perform segmentation. Each word was assigned a category in order to enable the distinction of target words and additional carrier words. The categories consisted of the 97 target words of the PLAKSS test plus an additional category "carrier word" for additional words that are not part of the test vocabulary. In the recordings of the 26 children, 2574 (26×99) target words were possible. However, only 2052 of the target words are present in the transliteration. This is related to the fact that the test was presented in pictograms. Hence, many children used alternatives to describe the pictogram. Sometimes, children also failed in the identification of a pictogram. As the test is rather long especially for young children with speech disorders, the therapist did not insist on the correct realization for each pictogram. She also counted alternatives as correct in order to keep the child motivated throughout the test.

In the next step the ASR system was used to segment the CLP-1 data into words and phones. All carrier words were excluded in the subsequent processing. Another 136 target words could not be used because the automatic segmentation failed, i.e., the segmented word was shorter than 100 ms. Hence, 93.3% of the appearing target words could be successfully segmented.

At the end of the semiautomatic segmentation procedure, 1916 words and 7647 phones, all from the target words, were obtained. This corresponds to 74.4% of the 2574 possible target words.

B. Fully automatic segmentation procedure

For the CLP-2 data set the semiautomatic segmentation procedure was replaced by a fully automatic one. Since the

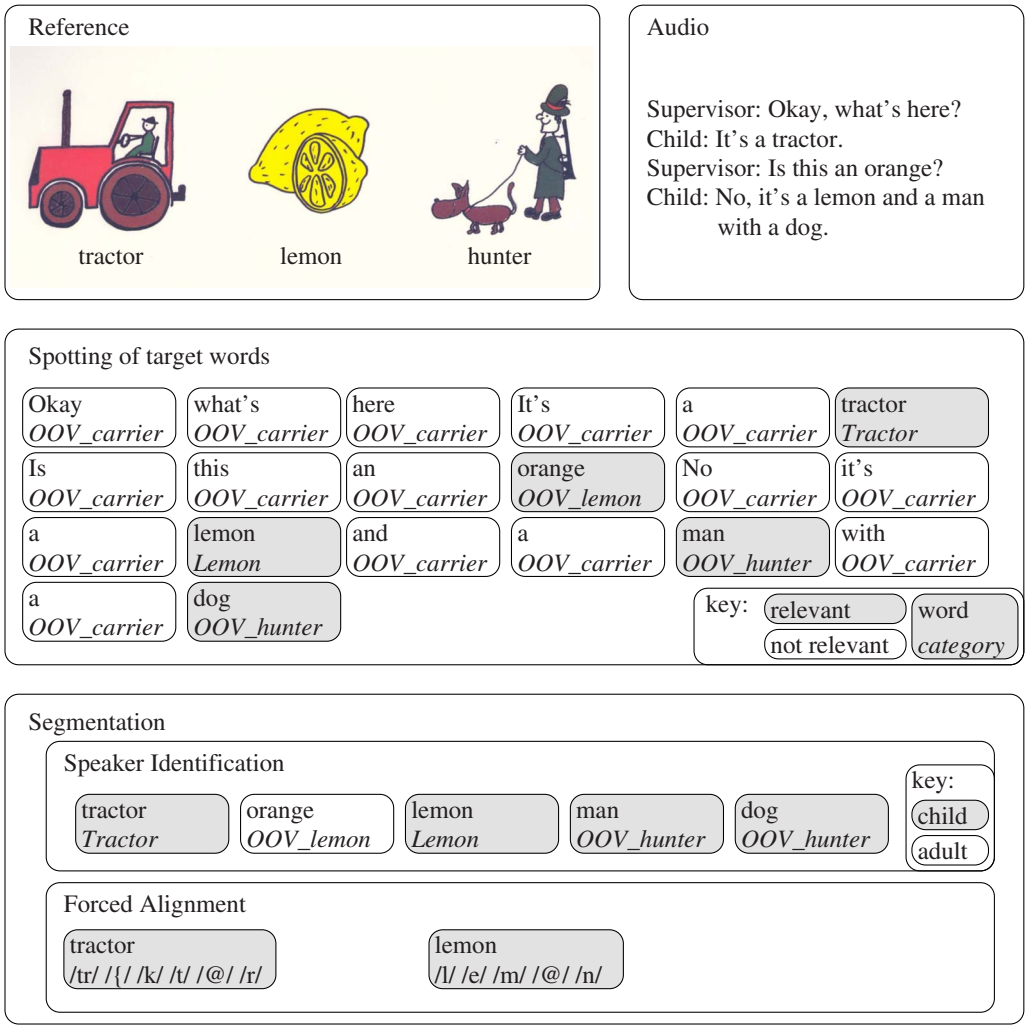


FIG. 3. (Color online) Diagram of the fully automatic word spotting and segmentation procedure: The audio data are processed by a speech recognition system. Its output is a chain of word and word category pairs. OOV words can also be detected. The category is used to identify the target words of the test. "Carrier words" are excluded from the subsequent processing. Alternatives of target words that still include the phonemic target are allowed. In the segmentation step, the speaker of each word (either the child or the therapist) is identified via energy thresholding. Finally, the successfully identified and categorized words are processed using forced alignment.

uttered word chain is not known *a priori* (cf. Sec. II), segmentation is much more difficult than in read speech, where the reference is known. First, candidates for target words have to be spotted. We do this using multiple ASR systems that are fused on word lattice level. Based on this recognition result, the target words are then extracted. Figure 3 shows a diagram of this processing, which is explained in more detail in the following.

In order to improve the segmentation, a speech recognition system with multiple trigram language models is used. The language models were created using the transliteration of the speech tests of 262 children with and without speech disorders. The categories of the language model were the 97 distinct words of the employed speech test, plus an additional category for words that appear in the "carrier sentences." In order to enable the recognition of misreadings, an out-of-vocabulary (OOV) word was added to each category.

Since the speech data were transliterated according to the acoustic realization of a child, the correspondence between the spoken words and the test words is not always clear. This is caused by the use of synonyms and pronunciation errors. However, every word of the transliteration has to be assigned to a category for the training of the language models. In order to solve this problem, an alignment was performed between the transliteration of 262 previously transliterated tests (of which the CLP-1 data are a subset) and the correct sequence of words with dynamic time warping.[31] In order to improve the alignment of pronunciation variants, the distance of substitutions was calculated according to the Levenshtein distance of the two words divided by the number of letters in the longer word. The procedure still has the problem that it is not capable of modeling variations in the sequence of the words that happen when a child names the words from right to left instead from left to right. Therefore, all found correspondences were checked manually according to their plausibility. Implausible correspondences were removed. So about 20 different alternatives of each word of the test were found.

In the transliteration of the 262 speech test sessions two tendencies could be seen: Some children use many carrier words while others use none at all. Furthermore, we built one turn-dependent and one turn-independent (using all of the

TABLE I. Articulation errors were annotated in the data by an experienced speech therapist according to the definition of Sell *et al.* (Ref. 5) in the group of 26 CLP children (CLP-1).

Speech disorder (criterion)	Description	Abbreviation
Hypernasality in vowels	Nasal air flow throughout the vowel	HN
Nasalized consonants	Air is emitted through the nose during the articulation of the consonants	NC
Pharyngealization	Tongue is shifted backward toward the pharynx during articulation	PB
Glottal articulation (laryngeal replacement)	Plosives are sucked to the larynx	LR
Weakened pressure consonants	Articulatory tension is diminished	WP

transliteration) model each. The turn-dependent models used only words that actually appeared in the transliteration of the processed turn to decrease the variability in the recognition. The turn-independent models were trained using all of the transliterations that were available and therefore allowed more variability. The segmentation is then performed using four language models for each turn: Two (one turn-dependent and one turn-independent) were trained on sentences with two or more carrier words per slide, and another two models with two or fewer carrier words. In preliminary experiments, trigram language models proved to yield the best recognition rates (RRs) in all four cases compared to language models with larger or smaller context.[27]

To estimate the probability of the OOV words, each word that occurred fewer than three times was used to train the OOV language model probabilities. The probabilities of the OOV words in the language model were estimated using the VOCSIM algorithm.[32] The acoustic realization of the OOV words is flat, i.e., it is assumed to be any sequence of the phonemes of the speech recognizer.

The recognition was performed for each turn using four different language models as described above. In order to obtain a single word chain, the four best word lattices plus the reference lattice, i.e., the actual object names, were merged using the recognizer output voting error reduction.[33–35]

In this manner, an improved recognized word chain is obtained. Preliminary experiments[27] were performed using the database with the 262 children. The data were split into training and a test set. All of the 26 children of the CLP-1 data were part of the test set. An increase in the word accuracy (WA) (cf. Sec. V A) of normal children speech from 64.7% to 74.5% was found. In the CLP speech data, this improvement was even more evident. The WA of −11.0% of the baseline system without any adaptation was pushed to 42.6%.

From the 3128 (32×99) target words that appear in the CLP-2 data, 2981 could be successfully extracted from the data. This corresponds to 94.0% of all target words. This percentage is much higher than in the semiautomatic case because the correct target names were shown below the pictogram. If we take a look at the successful segmentation rate of the semiautomatic system only, both are comparable (93.3% in the semiautomatic case).

IV. PERCEPTUAL EVALUATION OF THE SPEECH DATA

A speech therapist with many years of specialized experience thoroughly evaluated the CLP-1 data set. She differentiated all criteria, as listed in Table I. The therapist evaluated all target words that appeared in the transliteration by marking each affected phone.

Two other speech therapists examined the feature "nasal emission" as the most frequent error of the CLP-1 subset with implicit differentiation of nasalized vowels and NCs on phoneme level. They marked each phone either as "nasal" or "non-nasal." Due to the automatic segmentation procedure, only the target words of the PLAKSS test that could be segmented automatically were evaluated.

V. AUTOMATIC SPEECH DISORDER EVALUATION SYSTEM

The automatic evaluation system is divided into preprocessing, feature extraction, classification, and results and concludes with a decision for a specific class.[36] A scheme is shown in Fig. 4. The entire procedure is performed on the frame, phoneme, word, and speaker levels. On each of these levels different state-of-the-art features are computed.

On the frame level, mel frequency cepstrum coefficients (MFCCs) hold relevant information for the articulation. As features on the phoneme level, we extract teager-energy-profile (TEP) features as they have been shown to be relevant for the detection of nasality in vowels.[18] Furthermore, we compute pronunciation features on the phoneme level (*PronFexP*) as they were successfully applied to pronunciation scoring of non-native speech.[37] On the word level, the pronunciation features of Hacker *et al.*[38] have also been shown to be effective for the assessment of the pronunciation of second language learners. Also prosodic features (*ProsFeat*) may hold relevant information on the speech characteristics.[39] Hence, they were also included in the assessment procedure on the word level. On the speaker level, i.e., using all audio data of the speaker without segmentation, we included Sammon features[40] and the recognition accuracy of an ASR system[22] as both have been shown to be correlated to the speech intelligibility. Table II reports a summary of these features.

In our classification system we apply the concept of "late fusion,"[41] i.e., we train a classifier for each level. Com-

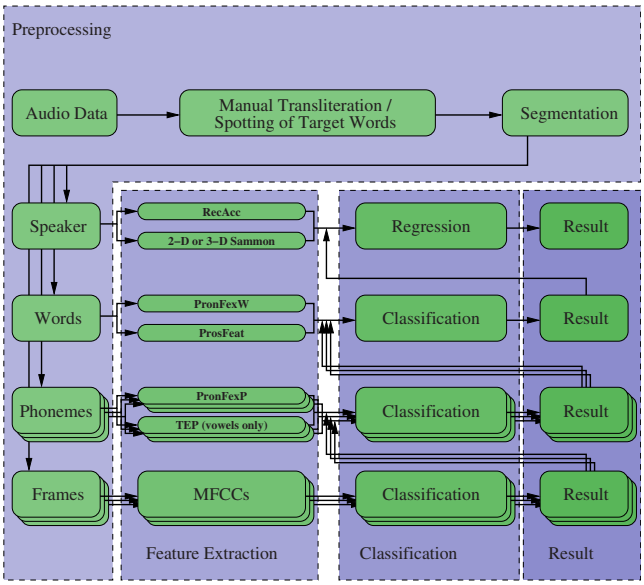


FIG. 4. (Color online) Experimental setup of the classification system: Right after the recording the preprocessing is performed. The data are transliterated manually or spotted automatically and segmented using forced alignment. Next, the feature extraction takes place on multiple levels. The features are supplied to a classifier that performs evaluation afterward. The output of each classifier is used as feature on the respective next evaluation level. On the frame and the phoneme levels, an independent classifier is trained for each phoneme. This is denoted as parallel arrows in the figure.

pared to other classification tasks, we do not have the option of "early fusion," i.e., concatenation of feature vectors, as this procedure would end in feature vectors of different length in vowels and consonants as some features can only be computed in vowels. Hence, we train different classifiers on the frame and phoneme levels for each phoneme. The output of each classifier is then used as feature on the respective next level. This is represented in Fig. 4 as multiple parallel arrows.

From the corresponding result of the lower level, features are computed and supplied to the higher level. These features include the mean, the maximum, the minimum, the standard deviation, the sum, and the product of the output probabilities. Furthermore, the absolute and relative frequencies of the decision for each class are regarded as features. Note that no information about the actual class membership is included in this process.

It is possible to compute evaluation results on all levels. We report these numbers to give an impression of the importance of the different feature groups on the respective levels. The main focus of this article, however, is the evaluation result on the speaker level.

For details on MFCCs, see, for example, Ref. 42. All other features are reported in detail in Secs. V A–V F. The section on the automatic speech disorder evaluation system is concluded by a description of the classifiers.

A. Recognition accuracy features

Good correlations between the intelligibility and the recognition accuracy have already been reported.[43] In our procedure we use the WR system described in Sec. III for processing the speech of the children. Then, the recognized word chain is compared to the reference, i.e., the target words, in order to determine the recognition accuracy.

In contrast to the segmentation procedure, we used a unigram language model to weigh the outcome of each word model. It was trained with the target words of the tests. Thus, the frequency of occurrence of each word in the used text was known to the recognizer. This enhances recognition results by including linguistic information. However, for our purpose it was also necessary to put more weight on the recognition of acoustic features. A comparison between unigram and zerogram language models was previously conducted.[44] It was shown that intelligibility can be predicted using WR accuracies computed by either zero- or unigram language models. The unigram, however, is computationally more efficient because it can be used to reduce the search space. The use of higher n-gram models was not beneficial in terms of correlation.[45]

For the evaluation of the recognized word chain, two measures are commonly used: the WR rate and the WA.

$$\mathrm{WR} = \frac{C}{R} \times 100\,\% \,.$$

TABLE II. Overview on the feature sets which are extracted on four different evaluation levels.

Label	Level	No.	Description	Reference
RecAcc	Speaker	2	Accuracy of the speech recognition (word correctness and accuracy)	22
2D Sammon Coordinates	Speaker	2	Coordinates on a 2D Sammon map	40
3D Sammon Coordinates	Speaker	3	Coordinates on a 3D Sammon map	40
ProsFeat	Word	37	Features based on the energy, the F_0, pauses, and duration to model the prosody of the speaker	39
PronFexW	Word	7	Pronunciation features (PronFex) to score the correctness of the current word	38
PronFexP	Phoneme	6	Features to score the correctness of the Pronunciation (PronFex) of the current phoneme	37
TEP	Phoneme	1	Teager energy profile to detect nasality in vowels	18
MFCCs	Frame	24	Mel frequency cepstrum coefficients	42

WR is computed as the percentage of correctly recognized words C and the number of reference words R. In addition,

$$\mathrm{WA} = \frac{C - I}{R} \times 100\%$$

weighs the number of wrongly inserted words I in this percentage. The WA punishes the insertion of additional words compared to the reference chain. Hence, it is known to be sensitive to carrier words.[46] The upper limit of both measures is 100%. The lower bound of the WR is 0% while the WA does not have a lower bound. It becomes negative, as soon as the recognizer inserts more wrong additional words than it actually recognizes correctly. This feature is used to support the assessment on the speaker level.

B. Sammon mapping

The speech data of each child are used to create speaker-dependent acoustic models. The adapted model coefficients are computed using MLLR adaptation of the speaker-independent model (cf. Sec. III). These adapted coefficients are then interpreted as a representation of the speaker with a fixed number of parameters, i.e., dimensions.

The Sammon transformation (ST) is a nonlinear method for mapping high dimensional data to a plane or a three dimensional (3D) space.[47] The ST uses the distances between the high dimensional data to find a lower dimensional representation—called map in this article. The ST preserves the topology of the original data, i.e., keeps the distance ratios between the low dimensional representations—called star here—as close as possible to the original distances. By doing so, the ST is cluster preserving. To ensure this, the function e_S is used as a measurement of the error of the resulting map [two dimensional (2D) case]:

$$e_S = s \sum_{p=1}^{N-1} \sum_{q=p+1}^{N} \frac{(\delta_{pq} - \theta_{pq})^2}{\delta_{pq}}, \tag{1}$$

with

$$\theta_{pq} = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2}. \tag{2}$$

δ_{pq} is the high dimensional distance between the high dimensional features p and q stored in a distance matrix $\boldsymbol{D}$, θ_{pq} is the Euclidean distance between the corresponding stars p and q in the map, and N is the total number of stars. For the computation of the high dimensional distance between two speech recognizers we use the Mahalanobis distance.[48,49] s is a scaling factor derived from the high dimensional distances:

$$s = \frac{1}{\sum_{p=1}^{N-1} \sum_{q=p+1}^{N} \delta_{pq}}. \tag{3}$$

The transformation is started with randomly initialized positions for the stars. Then the position of each star is optimized, using a conjugate gradient descent library.[50] This method is referred to as comprehensive space map of objective signal by Nagino and Shozakai.[51] A further advantage of the Sammon transform is that the derived coordinates can also yield further information on the intelligibility of the speaker.[52] In our experiments, e_S was 9% of the high dimensional distances. Hence, these features can be used on the speaker level for the assessment.

C. Prosodic features

The prosody module takes the output of our WR module in addition to the speech signal as input. In this case the time-alignment with the Viterbi algorithm of the recognizer and the information about the underlying phoneme classes (e.g., *long vowel*) are used by the prosody module.[53]

First, the prosody module extracts so-called base features from the speech signal. These are the energy, the fundamental frequency (F_0) after Bagshaw *et al.*,[54] and the voiced and unvoiced segments of the signal. In a second step, the actual prosodic features are computed in order to model the prosodic properties of the speech signal. For each word point, we extract 21 prosodic features. These features model F_0, energy, and duration. In addition, 16 global prosodic features for the whole utterance, i.e., slide, are calculated. They cover each of mean and standard deviation for jitter and shimmer,[55,56] the number, length, and maximum length each for voiced and unvoiced sections, the ratio of the numbers of voiced and unvoiced sections, the ratio of length of voiced sections to the length of the signal, and the same for unvoiced sections. The last global feature is the variance of the fundamental frequency F_0. Batliner *et al.*[53] presented a more detailed description of these features.

D. Pronunciation features

Articulation disorders result in pronunciation errors. Some of these errors concern confusion of phonemes. This is similar to the misarticulations that occur with speakers of a foreign language. Therefore, the investigation of methods that were developed for the scoring of non-native speech seems beneficial. Pronunciation features[57] are used to measure the progress in learning a foreign language.[58] In this work, we study these features' applicability to the detection of pathologic speech. More precisely, we only analyze a subset of these features that is based on phoneme confusion probabilities on a word level. To calculate these phoneme confusion features, we compare the result of the forced alignment with the Viterbi algorithm of every word to the result of a phoneme recognizer. The phoneme recognizer uses semi-continuous HMMs and a 4-g language model. It is based on MFCCs calculated every 10 ms with a frame size of 16 ms (cf. Sec. III). From the reference word and the recognized phoneme chain a confusion matrix $\boldsymbol{C}$ is built. It contains for every pair of phonemes a, b the probability that a was detected by the recognizer when there should be b according to the forced alignment

$$c_{ab} = P(a|b), \tag{4}$$

where c_{ab} is the corresponding entry of matrix $\boldsymbol{C}$. From the training set, we calculate two confusion matrices: one for the pathologic speech data and one for the normal data. From these framewise results, we calculate the following features for the phoneme level.[59]

- Goodness of pronunciation:[60] Score computed from the framewise score of a forced alignment and the likelihood obtained by a phoneme recognizer that was trained with normal speech. In non-native speech, the likelihood is known to drop in mispronounced phonemes. We expect the same for pathologic speech.
- Duration score: Likelihood of the observed phoneme duration given the duration distribution observed in normal speakers.
- Acoustic score: Posterior probability of the speech recognizer for the current phoneme.
- Actual duration: Observed duration.
- Expected duration: Mean value of the duration distribution observed in normal speakers.
- Confidence score Q:

$$Q = \frac{P_{\text{pathologic}}(a|b)}{P_{\text{normal}}(a|b)}. \tag{5}$$

For the word level, the following features are extracted:[37]

- PC1: Mean of Q,
- PC2: Maximum of Q,
- PC3: Minimum of Q,
- PC4: Variance of Q,
- PC5: Median of Q,
- A1: Phoneme correctness, and
- A2: Confidence score of the recognized word computed by the speech recognizer (cf. Sec. III).

E. Teager energy profiles

The teager energy operator (TEO) is a heuristic approach of pronunciation feature extraction. The teager operator[61] has been applied to detect nasality in sustained vowels and consonant-vowel-consonant combinations.[18] The TEO is defined as

$$\psi[f(n)] = [f(n)]^2 - f(n+1)f(n-1), \tag{6}$$

where $f(n)$ denotes the time-domain audio signal. The TEO's output is called the TEP.

The TEP can be used to detect hypernasal speech because it is sensitive to multicomponent signals.[13,18] When normal speech is low-pass-filtered in a way that the maximum frequency f_{lowpass} is somewhere between the first and the second formant, the resulting signal mainly consists of the first formant. However, the same procedure with hypernasal speech results in a multicomponent signal due to the antiformant.

In order to get a reference signal that contains only the first formant a second signal is computed with a band-pass filter around the first formant. The bandpass filter covers the frequency range ± 100 Hz around the first formant. For both signals the TEP is computed and compared. We measure that difference with the correlation coefficient between both TEPs. The values with the best results for f_{lowpass} were determined experimentally.[62]

F. Classification

For the classification various algorithms as provided in the Waikato environment for knowledge analysis[63] were employed. The following classifiers were used.

- *OneR*. The classifier divides the numeric features—often called attributes in machine learning—into intervals that contain only observations—also called instances—of one class. In order to prevent overfitting, mixed intervals are also allowed. However, each interval must hold at least a given number of instances in the training data. Then a decision rule for classification is created for each attribute. At the end of the training procedure, the attribute is selected for the classification that has the highest accuracy on the training set.[64]
- *DecisionStump*. DecisionStumps are commonly used in ensemble training techniques such as boosting. The classifier selects one attribute and a threshold or decision value to perform the classification. Selection is performed with correlation in the numeric case and entropy in the nominal case. Then, the selection value with the highest classification rate on the training set is determined.
- *LDA-Classifier*. The linear discriminant analysis (LDA)-Classifier is also called "ClassificationViaRegression."[65] It basically determines a LDA feature transformation matrix and reduces the dimension to 1. Then, a simple threshold can be chosen to perform the classification. Again, the threshold is determined on the training set according to the best classification rate.
- *NaiveBayes*. The naïve Bayes classifier is trained according to the Bayes' decision rule.[36] A unimodal Gaussian mixture is often chosen as a probability density function.[66] This classifier is equivalent to a Gaussian mixture model (GMM) classifier with just one Gaussian distribution with equal prior probabilities.
- *J48*. The J48 is an implementation of a C4.5 decision tree.[67] In order to build a C4.5 decision tree, all instances in the data set are used to create a set of rules. Later on, the rules are pruned in order to reduce their number. Subsequently, a tree is generated that holds one simple decision rule concerning only one attribute, i.e., a DecisionStump in every node. At the leaves of the tree a class label is assigned. Classification is then performed starting from the tree root and following a path according to rules in the node. At the end of the classification, a leaf is reached that assigns the class to the observation.
- *PART*. In order to modify the rules for a decision tree, two dominant approaches exist. The first one is eliminating rules like the J48 tree does. The second one extends rules by replacing one or multiple rules by a better more refined rule. PART generates partial trees using both approaches and merges them later on. This method is much faster in training compared to J48 while having a similar or even better recognition accuracy.[68]
- *RandomForest*. This kind of classifier is composed of multiple trees that are created randomly. For each tree a random subset of the training data is chosen. Then, a random subset of attributes is selected to be used in the tree. At each node, features are picked at random to determine the

rule of the actual node. The rule that creates the best split for the current subset is computed. Such a random tree may not be pruned. A set of a random number of trees is then fused into a random forest.[69]

- *SVM*. A support vector machine classifier[70] tries to find a surface that separates two classes from each other. Therefore, it is not necessary to remember all observations in the training set. Only a small number of observations is really important for the classification task. More specifically, only those feature vectors that are close to the class boundary are important for the decision. Although only two-class problems are considered in our work, the SVMs can easily be adapted to multiple classes by training an individual classifier for each class. In this manner a set of two-class problems "one against all others" is created.
- *AdaBoost*. Boosting[71] is a common procedure for enhancing simple classifiers. The idea of boosting is to join many weak classifiers to one single strong classifier. This is achieved by training in several iterations. In each iteration, the data are reweighted. Previously wrongly classified instances get a higher weight while correctly classified ones get a reduced weight adapting the classifiers to the misclassified instances.

We tested each of the classifiers at every level. The use of different classifiers on different levels was also allowed, but it was not permitted on the same level, e.g., different classifiers for different phonemes. The prior distribution of the classes, e.g., the probability of a word to be marked as "hypernasal," was not changed for the classification task since we wanted to keep the experiments as realistic as possible.

VI. EXPERIMENTAL RESULTS

In the following we present the results obtained by perceptual, semiautomatic, and automatic processing. The evaluation units are frames, phonemes, and words on the respective level. All evaluation measures are computed from the confusion matrix:

$$\begin{bmatrix} TP_a & FN_b \\ FN_a & TP_b \end{bmatrix}. \tag{7}$$

TP_a is the number of true positive classifications or the observable agreement for class Ω_a, i.e., that is unit as pathologic. FN_a is the number of false negatives, i.e., that the unit is wrongly assigned to the opposite class. This can also be referred to as the observable disagreement of class Ω_a. TP_b and FN_b are defined analogously.

The mostly used measure in classification tasks is the RR defined as

$$\mathrm{RR} = \frac{TP_a + TP_b}{N} \times 100\,\%\,, \tag{8}$$

where $N = TP_a + FN_a + TP_b + FN_b$. Furthermore, we introduce the classwise averaged RR (CL)

TABLE III. Results of the perceptual evaluation of the CLP-1 database.

	No. of affected		
Speech disorder	Phones	Words	Children
Hypernasality in vowels (HN)	49	49	4
Nasalized consonants (NCs)	329	329	15
Pharyngealization (PB)	34	33	7
Glottal articulation (LR)	32	31	4
Weakened pressure consonants (WPs)	105	105	14
Total number of units	7647	1916	26

$$\mathrm{CL} = \frac{1}{2} \times \left(\frac{TP_a}{TP_a + FN_a} + \frac{TP_b}{TP_b + FN_b} \right). \tag{9}$$

The CL is also often referred to as the unweighted average recall. The recall is defined as the number of true positives divided by the number of true positives and false negatives and is, therefore, equal to the definition of the sensitivity. Furthermore, we report the multirater-κ after Davies and Fleiss.[72]

The frame, phoneme, and word level results, however, are only intermediate results. The main focus of this article is the speaker level assessment. For each speaker we compute the percentage of pathologic words. Furthermore, we compute the percentage of detected words. Then, we measure the agreement using Pearson's correlation coefficient.[73]

A. Results of the perceptual evaluation

Table III reports the number of phones, words, and children that were affected by each of the five disorders according to the speech therapist's evaluation in the CLP-1 data set. The number of words is almost the same as the number of misarticulated phonemes since a single articulation error within a word meant that the whole word was counted as disordered. Only two words in the data set contained the same type of articulation error twice, i.e., 34 phonemes in 33 words with PB and 32 phonemes in 31 words with LR were annotated (cf. Table III). The last column shows the number of children who were affected by different disorders. While HN, PB, and LR appear in only few children, WP and NC appear in more than half of the children.

Table IV shows the agreement of the both raters of the CLP-2 data set. In the perceptual evaluation of the CLP-2 subset, the agreement of both raters was moderate. Only 127 words were marked as nasal emission by both raters. 2499 of the 2981 words were not marked as nasal emission. This

TABLE IV. Confusion matrix of the ratings by the two speech therapists for the criterion "nasal emission" on the CLP-2 database on the word level: Both raters marked 127 words as "nasal" and 2499 as "non-nasal." For 355 words the raters disagreed (κ=0.352).

Nasal emission	Nasal (rater 1)	Non-nasal (rater 1)
Nasal (rater 2)	127	203
Non-nasal (rater 2)	152	2499

TABLE V. Overview on the results of the pronunciation assessment on the frame, phoneme, word, and speaker levels for the CLP-1 data: All reported correlations (r) on the speaker level are significant at $p<0.01$.

	Semiautomatic evaluation									
	Frame			Phoneme			Word			
Criterion	CL (%)	RR (%)	κ	CL (%)	RR (%)	κ	CL (%)	RR (%)	κ	Speaker r
HN	56.8	99.0	0.564	62.9	99.0	0.627	60.6	96.9	0.596	0.89
NC	62.0	94.2	0.606	68.5	95.6	0.671	63.6	82.5	0.576	0.85
LR	59.8	99.6	0.597	69.5	99.6	0.694	63.8	98.2	0.632	0.81
PB	66.0	99.1	0.659	76.9	99.6	0.768	67.9	98.2	0.673	0.70
WP	71.1	97.8	0.708	71.1	97.8	0.707	75.8	97.8	0.745	0.82

corresponds to a true positive rate of the human rater 1 of 45.5% at a false positive rate of 7.5% taking rater 2 as the reference. Rater 2 had a true positive rate of 61.5% with a false negative rate of 5.7% taking rater 1 as the reference. κ values were 0.342 on the frame level, 0.313 on the phoneme level, and 0.352 on the word level.

In order to compare the automatic system with the perceptual evaluation, we computed both measures for each of the human raters (cf. Table IV). RR is the same for both raters, i.e., the percentage of observations where both raters agree:

$$\mathrm{RR} = \frac{127 + 2499}{2981} = 88.1\ \% .$$

CL is different for each rater. For rater 1, rater 2 is the reference:

$$\mathrm{CL(rater\ 1)} = \left(\frac{127}{127 + 152} + \frac{2499}{2499 + 203} \right) \Big/ 2 = 69.0\ \% ,$$

and for rater 2 rater 1 becomes the reference:

$$\mathrm{CL(rater\ 2)} = \left(\frac{127}{127 + 203} + \frac{2499}{2499 + 152} \right) \Big/ 2 = 66.4\ \% .$$

Correlation on the speaker level showed good consistency. When the percentages of marked words per speaker of both raters were compared a correlation of 0.80 was obtained.

B. Results of the automatic evaluation

All automatic evaluation experiments on the frame, phoneme, and word levels were conducted as leave-one-speaker-out evaluation, i.e., the training of the classifiers was performed with all but one speaker who was then employed as test speaker. This process was performed for all speakers.

To obtain a reference to build the automatic system, the label nasal emission is assigned if both raters agreed on their decision on the label in the CLP-2 data. Everything else was considered to be non-nasal. As reference on the speaker level, the percentage of marked words was chosen.

As reported in Table V very high values are reached for RR for the CLP-1 data set. This, however, is related to the unbalanced test sets: Most samples in the test set are not pathologic. Hence, classification of all samples to the class "normal" already yields high RRs. In order to optimize the CL rate, the training samples were weighted to form a balanced training set. The CL shows that the accuracy is moderate in most cases for these two class problems. The κ values are lower than in the semiautomatic case ($\kappa \approx 0.45$). This is related to the moderate agreement of the two raters (cf. Table IV; $\kappa \approx 0.35$), which is used in the multirater-κ computation. If we regard only the reference which was actually shown to the classifier in the training, κ lies in the same range as in the semiautomatic case ($\kappa \approx 0.6$).

On the speaker level, the features RecAcc and 2D or 3D Sammon coordinates (cf. Table II) were added to the evaluation procedure. Significance tests revealed that all reported correlations are highly significant with $p<0.01$. Except for WP, the result of the semiautomatic system achieves correlations above 0.81 for the phonetic disorders.

On the CLP-2 data, only the criterion nasal emission was evaluated (cf. Table VI). Again, high RRs were found in all classification experiments. As in the CLP-1 data, this is related to the bias in the distribution of the classes. The CL on the frame level is lower than the CLs for HN and NC in the CLP-1 data. On the phoneme level, this difference is already compensated. The CL of 64.8% is in between the recognition results of the HN and NC criteria. The same result can be observed on the word level. On the speaker level, a high correlation to the perceptual evaluation of the

TABLE VI. Overview on the results of the fully automatic pronunciation assessment on the frame, phoneme, and word levels for the CLP-2 data. The reported κ values are computed using the multirater-κ. The κ values in parentheses are computed using just the reference and the outcome of the automatic system. The correlation on the speaker level was significant with $r=0.81$ and $p<0.01$.

	Fully automatic evaluation		
Measure	Frame	Phoneme	Word
CL	52.6%	64.8%	62.1%
RR	98.8%	98.8%	94.0%
κ	0.431 (0.521)	0.478 (0.645)	0.482 (0.605)

TABLE VII. Detailed results for the different features on the frame, phoneme, and word levels for the CLP-1 and the CLP-2 data. If more than one rater was available (CLP-2 data only), κ values in parentheses report the agreement between the automatic system and the reference only.

Disorder	Feature	Level	CL (%)	RR (%)	κ
HN	MFCCs	**Frame**	**56.8**	**99.0**	**0.564**
HN	MFCCs	Phoneme	56.9	99.0	0.566
HN	TEP	Phoneme	59.2	97.7	0.589
HN	MFCCs+TEP	**Phoneme**	**62.9**	**99.0**	**0.627**
HN	MFCCs+TEP+PronFexP	Phoneme	60.6	98.7	0.603
HN	MFCCs	Word	52.3	96.9	0.511
HN	MFCCs+TEP	Word	57.7	95.8	0.566
HN	MFCCs+TEP+PronFex	**Word**	**60.6**	**96.9**	**0.596**
HN	MFCCs+TEP+PronFex+ProsFeat	Word	56.8	62.0	0.557
NC	MFCCs	**Frame**	**62.0**	**94.2**	**0.606**
NC	MFCCs	Phoneme	66.7	94.6	0.653
NC	PronFexP	Phoneme	67.5	91.5	0.661
NC	MFCCs+PronFexP	**Phoneme**	**68.5**	**95.6**	**0.671**
NC	MFCCs	**Word**	**63.6**	**82.5**	**0.576**
NC	MFCCs+PronFex+ProsFeat	Word	58.4	62.9	0.515
LR	MFCCs	**Frame**	**59.8**	**99.6**	**0.597**
LR	MFCCs	**Phoneme**	**69.5**	**99.6**	**0.694**
LR	MFCCs+PronFexP	Phoneme	65.3	92.6	0.652
LR	MFCCs	**Word**	**63.8**	**98.2**	**0.632**
LR	MFCCs+PronFex	Word	60.0	81.1	0.594
LR	MFCCs+PronFex+ProsFeat	Word	57.7	72.6	0.570
PB	MFCCs	**Frame**	**66.0**	**99.1**	**0.659**
PB	MFCCs	Phoneme	66.7	99.6	0.666
PB	MFCCs+PronFexP	**Phoneme**	**76.9**	**99.6**	**0.768**
PB	MFCCs	Word	59.8	98.2	0.591
PB	MFCCs+PronFex	**Word**	**67.9**	**98.2**	**0.673**
WP	MFCCs	**Frame**	**71.1**	**97.8**	**0.708**
WP	MFCCs	**Phoneme**	**71.1**	**97.8**	**0.707**
WP	MFCCs+PronFexP	Phoneme	71.0	88.5	0.706
WP	MFCCs	Word	66.1	97.8	0.642
WP	MFCCs+PronFex	Word	67.7	70.7	0.659
WP	MFCCs+PronFex+ProsFeat	**Word**	**75.8**	**97.8**	**0.745**
Nasalization	MFCCs	**Frame**	**52.6**	**98.8**	**0.431** (**0.521**)
Nasalization	MFCCs	Phoneme	62.4	98.7	0.466 (0.620)
Nasalization	MFCCs+TEP	Phoneme	62.0	98.7	0.464 (0.616)
Nasalization	MFCCs+TEP+PronFexP	**Phoneme**	**64.8**	**98.8**	**0.478** (**0.645**)
Nasalization	MFCCs	**Word**	**62.1**	**94.0**	**0.482** (**0.605**)
Nasalization	MFCCs+TEP	Word	60.2	81.8	0.472 (0.585)
Nasalization	MFCCs+TEP+Pronfex	Word	59.7	68.6	0.469 (0.580)

human raters of 0.81 is achieved. This is in the same range as the inter-rater correlation. No significant difference in the regression between nasality in vowels and the nasality in consonants was found on the speaker level ($p > 0.05$).

Table VII reports a detailed overview on the classification performance of different combinations of features. The best combinations are printed in boldface.

VII. DISCUSSION

As shown on the CLP-1 data, the system detects speech disorders on the speaker level as well as an expert. The correlations between the automatic system and the human expert for the different articulation disorders were mostly in the same range. Except for WP all correlation coefficients do not differ significantly from the best correlation of 0.89 ($p > 0.05$).

The speaker level evaluation of a fully automatic system performs comparably to two experienced listeners. The proposed system was tested for nasal emissions on the CLP-2 data. We decided for nasal emissions since they are the most characteristic and frequently occurring feature of speech of children with CLP. For our classification system, the differentiation of HN in vowels and HN in consonants does not play a significant role on the speaker level. As we train dif-

ferent classifiers for each phoneme this difference is compensated by the structure of our evaluation system on the higher evaluation levels.

For both experiments, the databases were suitable for this task since both contained a sufficient amount of normal and disordered speech data. The distribution of the classes normal and "disordered" in the test data was not adjusted, so as to create an evaluation task as realistic as possible.

Although the agreement between the human raters on the frame, phoneme, and word levels was moderate, we decided to use all data to train and test the classifiers. Selection of clear prototypical cases could, of course, improve the classification performance, as shown by Seppi *et al.*[74] However, as soon as the classifier is presented less prototypical test data, the classification performance drops significantly. Since we want to create a system that is employed in clinical routine use, we also need nonprototypical data.

In the semiautomatic system 93.3% of the target words that actually appeared in the audio data were usable for the subsequent processing. The fully automatic preprocessing procedure was able to replace this step completely. With the correct target words shown on the screen, 94.0% of them could be extracted successfully.

The system employs many state-of-the-art features and algorithms that are commonly used in pronunciation scoring of second language learners. It was shown that they also work for the evaluation of disordered speech.

Surprisingly, MFCCs alone yield high RR. We relate this to the fact that MFCCs model well human perception of speech in general. Hence, the effect of articulation disorders can also be seen in the MFCCs.

The features for transferring the classification output from one level to the next higher level are very useful. From the frame to phoneme levels, the recognition virtually always increased. On the word level, the phoneme level features also contributed to the recognition.

Combination of multiple features is beneficial on all evaluation levels, especially the pronunciation features in all articulation disorders and the TEP in the disorders concerning nasal emissions. Hence, the pronunciation features can not only model the pronunciation errors by non-natives but also articulation disorders in children. The TEP, which was previously only used in vowels and consonant-vowel combinations, showed to be applicable to connected speech as well. On the speaker level, RecAcc and Sammon coordinates increased the correlation to the perceptive evaluation. Prosodic features performed weakly in general. In most cases they did not contribute to any improvement. We relate this to the fact that the PLAKSS test is based on individual words and therefore induces only little prosody.

The employed classification toolbox provides state-of-the-art classifiers and methods for their combination. In general the tree-based classifiers, the SVMs, but also the DecisionStumps and NaïveBayes Classifiers combined with AdaBoost yielded the best performance.

On the frame and phoneme levels, CLs of up to 71.1% were reached on the CLP-1 data. On the word level the best CL was 75.8%. This is comparable to other studies concerning pronunciation scoring.[38,57,75] However, we consider these rates only as intermediate results indicative of the capabilities of the classification. Although there were errors, the classification errors are systematic. In contrast to commonly used perceptual evaluation by human listeners, results are not biased by individual experience. An automatic system therewith could provide different cleft centers with a standardized detection method for speech disorders. The classification on the word level with 75.8% CL is sufficient for a good quantification of all five disorders on the speaker level, as can be seen in the high and significant correlations (0.70–0.89). The classification system shows errors but they are consistent, i.e., the number of additional instances that are classified as disordered is similar in all speakers. The percentage of disordered events can be predicted reliably by regression.

The lowest correlation was found to be 0.70 for PB while the best correlation was 0.89 for HN in the CLP-1 data. All correlations were highly significant with $p<0.01$. In previous studies we found inter-rater correlations in the same range between human experts for the same evaluation tasks.[20]

On the CLP-2 data, CLs and RRs for experiments on the frame, phoneme, and word levels were comparable to the semiautomatic case. κ, however, was reduced. This is caused by the moderate inter-rater agreement between the two human raters ($\kappa \approx 0.35$), which is also included in the computation of the multirater-κ. Hence, κ dropped from approximately 0.6 to 0.45. As we focus on the automatic evaluation and the performance of the automatic system in this article, it is also valid to regard only the reference that was actually shown to the classifier. In this manner we simulate a single rater. Then, κ values are comparable to the semiautomatic, single-rater case ($\kappa \approx 0.6$), i.e., in both cases the classifiers do their task and learn the shown reference in a comparable manner.

The evaluation on a speaker level also had a high and significant correlation of 0.81 ($p<0.01$). The human listeners had an inter-rater correlation of 0.80, which is enough to quantify speech disorders on a speaker level sufficiently. There is no significant difference between human-human and the human-machine correlations ($p>0.05$). Hence, the evaluation of the fully automatic system is at an expert's level. The intrarater correlation of the automatic system is 1 since it always quantifies the same input with the same degree of nasal emissions. The automatic system can be regarded as a fast and reliable way to evaluate nasal emissions in speech of children with CLP at an expert's level. Of course, the application on other phonetic disorders will be realized. Hence, the fully automatic system is suitable for clinical use.

VIII. SUMMARY

This paper presents the first automatic evaluation system for distinct articulation disorders in connected speech. The system has been evaluated on articulation disorders of children with CLP with different extent and characteristics of phonetic disorders. Since the usually applied perceptual evaluation of these disorders requires a lot of time and manpower, there is a need for quick and objective automatic

evaluation. To investigate the evaluation by an automatic system, two experiments with articulation disorders were conducted. On one data set (CLP-1), a test for five characteristic articulation disorders was performed by an experienced speech therapist to show to show that the system is able to detect different articulation disorders. On the second database (CLP-2), the evaluation was performed with a fully automatic system without any additional human effort.

On the frame, phoneme, and word levels, the performance is moderate. On the speaker level, however, the system shows good correlations to the commonly used perceptual evaluation by expert listeners. The correlation between the system and the perceptual evaluation was in the same range as the inter-rater correlation of experienced speech therapists. Thus, the system will facilitate the clinical and scientific evaluation of speech disorders.

ACKNOWLEDGMENTS

This work was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft DFG) under Grant No. SCHU2320/1-1. We thank Dr. Ulrike Wohlleben, Andrea Schädel, and Dorothee Großmann for the expert's annotation of the data. Furthermore, we would like to thank the anonymous reviewers of this article and the editor for their through comments on our work.

[1]R. Ruben, "Redefining the survival of the fittest: Communication disorders in the 21st century," Laryngoscope **110**, 241–245 (2000).
[2]B. Eppley, J. van Aalst, A. Robey, R. Havlik, and M. Sadove, "The spectrum of orofacial clefting," Plast. Reconstr. Surg. **115**, 101e–114e (2005).
[3]M. Tolarova and J. Cervenka, "Classification and birth prevalence of orofacial clefts," Am. J. Med. Genet. **75**, 126–137 (1998).
[4]H. Kawamoto, "Rare craniofacial clefts," in *Plastic Surgery*, edited by J. C. McCarthy (Saunders, Philadelphia, PA, 1990), Vol. **4**.
[5]D. Sell, P. Grunwell, S. Mildinhall, T. Murphy, T. Cornish, D. Bearn, W. Shaw, J. Murray, A. Williams, and J. Sandy, "Cleft lip and palate care in the United Kingdom—The Clinical Standards Advisory Group (CSAG) study. Part 3: Speech outcomes," Cleft Palate Craniofac J. **38**, 30–37 (2001).
[6]J. Karling, O. Larson, R. Leanderson, and G. Henningsson, "Speech in unilateral and bilateral cleft palate patients from Stockholm," Cleft Palate Craniofac J. **30**, 73–77 (1993).
[7]K. Van Lierde, M. D. Bodt, J. V. Borsel, F. Wuyts, and P. V. Cauwenberge, "Effect of cleft type on overall speech intelligibility and resonance," Folia Phoniatr Logop **54**, 158–168 (2002).
[8]K. Van Lierde, M. D. Bodt, I. Baetens, V. Schrauwen, and P. V. Cauwenberge, "Outcome of treatment regarding articulation, resonance and voice in Flemish adults with unilateral and bilateral cleft palate," Folia Phoniatr Logop **55**, 80–90 (2003).
[9]M. Hardin, D.-R. Van Demark, H. Morris, and M. Payne, "Correspondence between nasalance scores and listener judgments of hypernasality and hyponasality," Cleft Palate Craniofac J. **29**, 346–351 (1992).
[10]T. Pruthi and C. Y. Espy-Wilson, "Automatic classification of nasals and semivowels," in *ICPhS 2003-15th International Congress of Phonetic Sciences*, Barcelona, Spain (2003), pp. 3061–3064.
[11]T. Pruthi and C. Y. Espy-Wilson, "Acoustic parameters for automatic detection of nasal manner," Speech Commun. **43**, 225–239 (2004).
[12]T. Pruthi, C. Y. Espy-Wilson, and H. Brad, "Story, simulation and analysis of nasalized vowels based on magnetic resonance imaging data," J. Acoust. Soc. Am. **121**, 3858–3873 (2007).
[13]D. Cairns, J. Hansen, and J. Kaiser, "Recent advances in hypernasal speech detection using the nonlinear teager energy operator," in *Proceedings of the International Conference on Speech Communication and Technology (Interspeech)* (ISCA, Philadelphia, PA, 1996), Vol. **2**, pp. 780–783.
[14]R. Kataoka, K. Michi, K. Okabe, T. Miura, and H. Yoshida, "Spectral properties and quantitative evaluation of hypernasality in vowels," Cleft Palate Craniofac J. **33**, 43–50 (1996).
[15]B. Atal and M. Schroeder, "Predictive coding of speech signals," in *Proceedings of the Conference Communication and Processing* (1967), pp. 360–361.
[16]G. Fant, "Nasal sounds and nasalization," *Acoustic Theory of Speech Production* (Mouton, The Hague, The Netherlands, 1960).
[17]A. Zečević, "Ein sprachgestütztes Trainingssystem zur Evaluierung der Nasalität (A speech-supported training system for the evaluation of nasality)," Ph.D. thesis, University of Mannheim, Mannheim, Germany (2002).
[18]D. Cairns, J. Hansen, and J. Riski, "A noninvasive technique for detecting hypernasal speech using a nonlinear operator," IEEE Trans. Biomed. Eng. **43**, 35–45 (1996).
[19]K. Keuning, G. Wieneke, and P. Dejonckere, "The intrajudge reliability of the perceptual rating of cleft palate speech before and after pharyngeal flap surgery: The effect of judges and speech samples," Cleft Palate Craniofac J. **36**, 328–333 (1999).
[20]S. Paal, U. Reulbach, K. Strobel-Schwarthoff, E. Nkenke, and M. Schuster, "Evaluation of speech disorders in children with cleft lip and palate," J. Orofac. Orthop. **66**, 270–278 (2005).
[21]F. Wuyts, M. D. Bodt, G. Molenberghs, M. Remacle, L. Heylen, B. Millet, K. V. Lierde, J. Raes, and P. V. Heyning, "The dysphonia severity index: An objective measure of vocal quality based on a multiparameter approach," J. Speech Lang. Hear. Res. **43**, 796–809 (2000).
[22]A. Maier, C. Hacker, E. Nöth, E. Nkenke, T. Haderlein, F. Rosanowski, and M. Schuster, "Intelligibility of children with cleft lip and palate: Evaluation by speech recognition techniques," in *Proceedings of the International Conference on Pattern Recognition (ICPR)*, Hong Kong, China (2006), Vol. **4**, pp. 274–277.
[23]A. Maier, F. Hönig, C. Hacker, M. Schuster, and E. Nöth, "Automatic evaluation of characteristic speech disorders in children with cleft lip and palate," in Interspeech 2008-*Proceedings of the International Conference on Spoken Language Processing*, 11th International Conference on Spoken Language Processing, Brisbane, Australia (2008), pp. 1757–1760.
[24]A. Maier, T. Haderlein, U. Eysholdt, F. Rosanowski, A. Batliner, M. Schuster, and E. Nöth, "PEAKS—A system for the automatic evaluation of voice and speech disorders," Speech Commun. **51**, 425–437 (2009).
[25]A. Fox, "PLAKSS—Psycholinguistische Analyse kindlicher Sprechstörungen (Psycholinguistic analysis of children's speech disorders)," Swets and Zeitlinger, Frankfurt a.M., Germany, now available from Harcourt Test Services GmbH, Germany (2002).
[26]J. Wilpon and C. Jacobsen, "A study of speech recognition for children and the elderly," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Atlanta, GA (1996), Vol. **1**, pp. 349–352.
[27]T. Bocklet, "Optimization of a speech recognizer for medical studies on children in preschool and primary school age," Diplomarbeit, Chair of Pattern Recognition, University Erlangen-Nuremberg, Erlangen, Germany (2007).
[28]A. Maier, "Recognizer adaptation by acoustic model interpolation on a small training set from the target domain," Diplomarbeit, Chair of Pattern Recognition, University Erlangen-Nuremberg, Erlangen, Germany (2005).
[29]A. Maier, T. Haderlein, and E. Nöth, "Environmental adaptation with a small data set of the target domain," in *Lecture Notes in Artificial Intelligence, Ninth International Conference on Text, Speech and Dialogue (TSD)*, edited by P. Sojka, I. Kopeček, and K. Pala (Springer, Berlin, 2006), Vol. **4188**, pp. 431–437.
[30]A. Maier, *Speech Recognizer Adaptation* (VDM, Saarbrücken, 2008).
[31]*Time Warps, String Edits, and Macromolecules*, edited by D. Sankoff and J. Kruskal (Addison-Wesley, Reading, MA, 1983).
[32]F. Gallwitz, *Integrated Stochastic Models for Spontaneous Speech Recognition* (Logos, Berlin, 2002), Vol. **6**.
[33]J. Fiscus, "A post-processing system to yield reduced word error rates: Recogniser output voting error reduction," in *Proceedings of the IEEE ASRU Workshop*, Santa Barbara, CA (1997), pp. 347–352.
[34]A. Maier, C. Hacker, S. Steidl, E. Nöth, and H. Niemann, "Robust parallel speech recognition in multiple energy bands," in *Lecture Notes in Computer Science, Pattern Recognition, 27th DAGM Symposium*, Vienna, Austria, edited by G. Kropatsch, R. Sablatnig, and A. Hanbury (Springer, Berlin, 2005), Vol. **3663**, pp. 133–140.
[35]A. Maier, *Parallel Robust Speech Recognition* (VDM, Saarbrücken, 2008).
[36]H. Niemann, *Klassifikation von Mustern (Pattern Classification)*, 2nd ed. (Springer, Berlin, 2003), http://www5.informatik.uni-erlangen.de/Personen/niemann/klassifikation-von-mustern/m00links.html (Last viewed 02/12/2008).

[37]T. Cincarek, "Pronunciation scoring for non-native speech," Diplomarbeit, Chair of Pattern Recognition, University Erlangen-Nuremberg, Erlangen, Germany (2004).
[38]C. Hacker, T. Cincarek, A. Maier, A. Heßler, and E. Nöth, "Boosting of prosodic and pronunciation features to detect mispronunciations of non-native children," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (IEEE Computer Society Press, Mayi, HI, 2007), Vol. **4**, pp. 197–200.
[39]T. Haderlein, E. Nöth, M. Schuster, U. Eysholdt, and F. Rosanowski, "Evaluation of tracheoesophageal substitute voices using prosodic features," in *Proceedings of the Speech Prosody, Third International Conference*, edited by R. Hoffmann and H. Mixdorff (TUD-Press, Dresden, 2006), pp. 701–704.
[40]T. Haderlein, D. Zorn, S. Steidl, E. Nöth, M. Shozakai, and M. Schuster, "Visualization of voice disorders using the Sammon transform," in *Lecture Notes in Artificial Intelligence, Ninth International Conference on Text, Speech and Dialogue (TSD)*, edited by P. Sojka, I. Kopeček, and K. Pala (Springer, Berlin, 2006), Vol. **4188**, pp. 589–596.
[41]C. Snoek, M. Worring, and A. Smeulders, "Early versus late fusion in semantic video analysis," in *Proceedings of the 13th Annual ACM International Conference of Multimedia* (ACM, New York, 2005), pp. 399–402.
[42]S. Davis and P. Mermelstein, "Comparison of parametric representation for monosyllabic word recognition in continuously spoken sentences," IEEE Trans. Acoust., Speech, Signal Process. **28**, 357–366 (1980).
[43]M. Schuster, A. Maier, T. Haderlein, E. Nkenke, U. Wohlleben, F. Rosanowski, U. Eysholdt, and E. Nöth, "Evaluation of speech intelligibility for children with cleft lip and palate by automatic speech recognition," Int. J. Pediatr. Otorhinolaryngol. **70**, 1741–1747 (2006).
[44]K. Riedhammer, G. Stemmer, T. Haderlein, M. Schuster, F. Rosanowski, E. Nöth, and A. Maier, "Towards robust automatic evaluation of pathologic telephone speech," in *Proceedings of the Automatic Speech Recognition and Understanding Workshop (ASRU)* (IEEE Computer Society Press, Kyoto, Japan, 2007), pp. 717–722.
[45]A. Maier, T. Haderlein, F. Stelzle, E. N. E. Nkenke, F. Rosanowski, A. Schützenberger, and M. Schuster, "Automatic speech recognition systems for the evaluation of voice and speech disorders in head and neck cancer," EURASIP J. Audio, Speech, and Music Processing **2010**, In press.
[46]A. Maier, E. Nöth, A. Batliner, E. Nkenke, and M. Schuster, "Fully automatic assessment of speech of children with cleft lip and palate," Informatica **30**, 477–482 (2006).
[47]J. Sammon, "A nonlinear mapping for data structure analysis," IEEE Trans. Comput. **C-18**, 401–409 (1969).
[48]P. Mahalanobis, "On the generalised distance in statistics," in Proceedings of the National Institute of Science of India **12**, 49–55 (1936).
[49]M. Shozakai and G. Nagino, "Analysis of speaking styles by two-dimensional visualization of aggregate of acoustic models," in *Proceedings of the International Conference on Speech Communication and Technology (Interspeech)* (ISCA, Jeju Island, Korea, 2004), Vol. **1**, pp. 717–720.
[50]W. Naylor and B. Chapman, "WNLIB homepage," http://www.willnaylor.com/wnlib.html (Last viewed 07/20/2007).
[51]M. Nagino and G. Shozakai, "Building an effective corpus by using acoustic space visualization (cosmos) method," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (IEEE Computer Society Press, Philadelphia, PA, 2005), pp. 449–452.
[52]A. Maier, M. Schuster, U. Eysholdt, T. Haderlein, T. Cincarek, S. Steidl, A. Batliner, S. Wenhardt, and E. Nöth, "QMOS—A robust visualization method for speaker dependencies with different microphones," J. Pattern Recognition Research **4**, 32–51 (2009).
[53]A. Batliner, A. Buckow, H. Niemann, E. Nöth, and V. Warnke, "The prosody module," in *Verbmobil: Foundations of Speech-to-Speech Translation*, edited by W. Wahlster (Springer, New York, 2000), pp. 106–121.
[54]P. Bagshaw, S. Hiller, and M. Jack, "Enhanced pitch tracking and the processing of f0 contours for computer aided intonation teaching," in *Proceedings of the European Conference on Speech Communication and Technology (Eurospeech)* (ISCA, Berlin, 1993), pp. 1003–1006.
[55]R. Kompe, *Prosody in Speech Understanding Systems*, Lecture Notes in Artificial Intelligence Vol. **1307** (Springer, Berlin, 1997).
[56]S. Steidl, *Automatic Classification of Emotion-Related User States in Spontaneous Children's Speech* (Logos, Berlin, 2009).
[57]C. Hacker, T. Cincarek, R. Gruhn, S. Steidl, E. Nöth, and H. Niemann, "Pronunciation feature extraction," in *Lecture Notes in Computer Science, Pattern Recognition, 27th DAGM Symposium*, Vienna, Austria, edited by G. Kropatsch, R. Sablatnig, and A. Hanbury (Springer, Berlin, 2005), Vol. **3663**, pp. 141–148.
[58]T. Cincarek, R. Gruhn, C. Hacker, E. Nöth, and S. Nakamura, "Gaikokugohatsuon no jidouhyoutei to yomiayamatta tango no jidoukenshutsu (Automatic evaluation of foreign language pronunciation and automatic recognition of reading errors in vocabulary)," in *Proceedings of the Acoustical Society of Japan* (2004), pp. 165–166.
[59]T. Cincarek, R. Gruhn, C. Hacker, E. Nöth, and S. Nakamura, "Automatic pronunciation scoring of words and sentences independent from the non-native's first language," Comput. Speech Lang. **23**, 65–99 (2009).
[60]S. Witt and S. Young, "Phone-level pronunciation scoring and assessment for interactive language learning," Speech Commun. **30**, 95–108 (2000).
[61]H. Teager and S. Teager, "Evidence for nonlinear production mechanisms in the vocal tract," in *Speech Production and Speech Modelling* (1990), pp. 241–261.
[62]A. Reuß, "Analysis of speech disorders in children with cleft lip and palate on phoneme and word level," Studienarbeit, Chair of Pattern Recognition, University Erlangen-Nuremberg, Erlangen, Germany (2007).
[63]I. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. (Kaufmann, San Fransisco, CA, 2005).
[64]R. Holte, "Very simple classification rules perform well on most commonly used datasets," Mach. Learn. **11**, 63–91 (1993).
[65]E. Frank, Y. Wang, S. Inglis, G. Holmes, and I. Witten, "Using model trees for classification," Mach. Learn. **32**, 63–76 (1998).
[66]G. H. John and P. Langley, "Estimating continuous distributions in Bayesian classifiers," in *11th Conference on Uncertainty in Artificial Intelligence* (Kaufmann, San Mateo, CA, 1995), pp. 338–345.
[67]R. Quinlan, *C4.5: Programs for Machine Learning* (Kaufmann, San Mateo, CA, 1993).
[68]E. Frank and I. H. Witten, "Generating accurate rule sets without global optimization," in *15th International Conference on Machine Learning*, edited by J. Shavlik (Kaufmann, San Mateo, CA, 1998), pp. 144–151.
[69]L. Breiman, "Random forests," Mach. Learn. **45**, 5–32 (2001).
[70]B. Schölkopf, "Support vector learning," Ph.D. thesis, Technische Universität Berlin, Berlin, Germany (1997).
[71]Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *13th International Conference on Machine Learning* (Kaufmann, San Mateo, CA, 1996), pp. 148–156.
[72]M. Davies and J. Fleiss, "Measuring agreement for multinomial data," Biometrics **38**, 1047–1051 (1982).
[73]K. Pearson, "Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia," Philos. Trans. R. Soc. London **187**, 253–318 (1896).
[74]D. Seppi, A. Batliner, B. Schuller, S. Steidl, T. Vogt, J. Wagner, L. Devillers, L. Vidrascu, N. Amir, and V. Aharonson, "Patterns, prototypes, performance: Classifying emotional user states," in Interspeech 2008-*Proceedings of the International Conference on Spoken Language Processing*, 11th International Conference on Spoken Language Processing, Brisbane, Australia (2008), pp. 601–604.
[75]A. Neri, C. Cuchiarini, and C. Strik, "Feedback in computer assisted pronunciation training: Technology push or demand pull?," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (IEEE Computer Society Press, Orlando, FL, 2002), pp. 1209–1212.

Cross-dialectal variation in formant dynamics of American English vowels

Robert Allen Fox[a] and Ewa Jacewicz
Department of Speech and Hearing Science, The Ohio State University, Columbus, Ohio 43210-1002

(Received 31 March 2009; revised 30 July 2009; accepted 30 July 2009)

This study aims to characterize the nature of the dynamic spectral change in vowels in three distinct regional varieties of American English spoken in the Western North Carolina, in Central Ohio, and in Southern Wisconsin. The vowels /ɪ, ɛ, e, æ, aɪ/ were produced by 48 women for a total of 1920 utterances and were contained in words of the structure /bVts/ and /bVdz/ in sentences which elicited nonemphatic and emphatic vowels. Measurements made at the vowel target (i.e., the central 60% of the vowel) produced a set of acoustic parameters which included position and movement in the F1 by F2 space, vowel duration, amount of spectral change [measured as vector length (VL) and trajectory length (TL)], and spectral rate of change. Results revealed expected variation in formant dynamics as a function of phonetic factors (vowel emphasis and consonantal context). However, for each vowel and for each measure employed, dialect was a strong source of variation in vowel-inherent spectral change. In general, the dialect-specific nature and amount of spectral change can be characterized quite effectively by position and movement in the F1 by F2 space, vowel duration, TL (but not VL which underestimates formant movement), and spectral rate of change.
 [DOI: 10.1121/1.3212921]

PACS number(s): 43.70.Fq, 43.72.Ar [DAB] Pages: 2603–2618

I. INTRODUCTION

This study is an acoustic investigation into time-varying spectral features of the vowel target and the use of these features by different regional variants (dialects) of American English. In the long tradition of research on vowel acoustics, the vowel target has been regarded as the central section of the vowel which is relatively unaffected by surrounding consonants (e.g., Lehiste and Peterson, 1961; Lindblom, 1963). As such, formant frequencies measured at the vowel target are often considered to characterize most appropriately a particular vowel quality with the general assumption that this target implies some degree of invariance or vowel's "steady state." As the research progressed, however, it has been recognized that even the most monophthongal vowel is never truly "static" and usually exhibits a certain amount of vowel-inherent spectral change (e.g., Harrington and Cassidy, 1994; Nearey and Assmann, 1986). This spectral change is not contextually determined but is a systematic property of the vowel itself.

Naturally, the presence of dynamic spectral change in a vowel gave rise to investigation in the perceptual domain, inviting the question of how important are the dynamic cues to vowel identification. Certainly a substantial body of research has focused on determining which aspects of the acoustic signal contribute most to vowel identification: a relatively steady target or rapid consonantal transitions. For example, work by Strange and Jenkins (e.g., Jenkins *et al.*, 1983, 1994; Strange, 1987, 1989; Strange *et al.*, 1976, 1983) underscored the role of consonantal transitions proposing that a vowel can be reliably identified even if its "center" has been removed experimentally from the signal. On the other hand, studies by Nearey and colleagues (e.g., Hillenbrand and Gayvert, 1993; Hillenbrand and Nearey, 1999; Nearey and Assmann, 1986; Kewley-Port and Neel, 2006) demonstrated that listeners identified vowels with greater accuracy when the vowel-specific pattern of spectral change was preserved at the vowel's center. The results from these two lines of research suggest that neither the vowel target nor consonantal transitions alone are fully sufficient for vowel identification.

The complexity of acoustic variation in formant dynamics comes from several sources, including the vowel-specific nature of trajectory change in the formant space, consonantal context effects, emphatic stress, or broadly defined prosodic effects. The difficulty in characterizing the dynamic spectral changes throughout the course of the vowel lies in the fact that these changes occur in time and are subject to temporal variation in speech. That is, the amount of vowel-inherent spectral change may vary with vowel duration, which is also affected systematically by consonantal and prosodic contexts as well as by variation in speech tempo. For example, consonantal context effects may persist throughout the vowel (including its target) if the vowel is sufficiently short. This leads to yet another complication, however, in that vowel target may not be easily defined as it is the entire formant trajectory that undergoes changes over time.

The presence of time-varying spectral features such as the amount of spectral change and spectral rate of change (roc) implies that the dynamic information includes cues from both spectral and temporal domains. As it has been shown and is argued below, both sets of cues are important to a better understanding of vowel dynamics and their use in speech communication. For example, according to the contextual model of articulation (Lindblom, 1963; Moon and

[a]Author to whom correspondence should be addressed. Electronic mail: fox.2@osu.edu

Lindblom, 1994), the amount of formant undershoot depends on the interaction between the phonetic context, vowel duration, and the spectral roc. The higher spectral roc is related to faster articulatory movements typically invoked to reach the formant target. An application of this model in concatenative synthesis showed that a better control of spectral roc improves the naturalness and intelligibility of synthesized utterances (Wouters and Macon, 2002b).

Dynamic variations at the vowel target do exist in naturally produced nominal monophthongs and have been found in selective acoustic studies of American, Canadian, and Australian English vowels (e.g., Andruski and Nearey, 1992; Hillenbrand *et al.*, 1995; Watson and Harrington, 1999). Yet, there is no acoustic evidence that vowel-inherent spectral change may actually vary systematically across geographic regions of the country and that the use of time-varying features may be a subject to regional variation. Sociolinguistic work on phonetics and vowel shifts, culminating in the *Atlas of North American English* (Labov *et al.*, 2006), has been primarily concerned with relative positions of vowels in order to create regional maps. While documenting vowel fronting, backing, lowering, raising, centralization, or mergers, formant measurements have been taken typically at one temporal location at the vowel target, referred to as the vowel nucleus. Although this procedure accounts for the regional differences in the positions of vowel nuclei, it does not address the issue of how vowels differ in the extent and nature of dynamic information which may also contribute to regional variation in American English. Crucially, information about how formant frequency changes in time is missing.

There is evidence that the duration of American English vowels varies significantly across regions in the United States (e.g., Clopper *et al.*, 2005; Jacewicz *et al.*, 2007) and so does speech tempo (e.g., Byrd, 1994; Jacewicz *et al.*, 2009). It can be expected that these temporal factors may have a profound effect on formant frequency change in the course of the vowel. As shown by Lindblom and colleagues, there is a complex interaction between vowel duration, consonantal context, and speaking style on formant frequency shifts so that both the position of the vowel in the acoustic space and its spectral dynamics will vary in predictable ways (e.g., Lindblom *et al.*, 2007; Moon and Lindblom, 1994). Expansion or reduction in the vowel space and degree of coarticulation with surrounding consonants are the most observable effects. However, in addition to these and other sources of phonetic variation such as emphatic stress and tempo, regional variation introduces yet another variable to be accounted for in characterizing the acoustic structure of vowels. Cross-dialectal differences in vowel duration, speech tempo, and the extent of formant change within the vowel pose a question of the importance of time-varying information in the differentiation of regional variants. As our understanding progresses, we can ask further questions such as to what extent do speakers of a specific dialect rely on the dynamic aspects of the acoustic signal in identifying vowels as coming from their own dialect.

The aim of the present study was to define the nature of the dynamic spectral change at the targets, that is, the central sections of selected vowels in three distinct regional varieties of American English spoken in the South (Western North Carolina), in the central region around Columbus, OH, and in the North (Southern Wisconsin). The vowels selected included a true diphthong (/aɪ/), a diphthongized long vowel (/e/), and three lax vowels which exhibited differences in the degree of their diphthongal and temporal properties (/ɪ, ɛ, æ/). The dynamic variations in the formants F1 and F2 were assessed in a set of acoustic measures: vector length (VL), trajectory length (TL), and the spectral roc. Two sources of phonetic variation were included which are known to affect systematically vowel duration: emphatic stress and consonantal context. The study sought to determine the extent to which the dialectal differences in spectral features are present in combination with expected spectral variation as a function of emphasis and context (Lindblom *et al.*, 2007).

II. METHODS

A. Speakers

Forty eight women aged 51–65 years participated in the study. They were born, raised, and spent most of their lives in one of three regions in the United States: 16 were from Western North Carolina (the Jackson County area), 16 were from Central Ohio (Columbus area), and 16 were from Southern Wisconsin (Madison area). Defined geographically, these participants created highly homogeneous samples of regional speech, deeply rooted in the regional dialect. According to the *Atlas of North American English* (Labov *et al.*, 2006), these dialects represent Inland South, the Midland, and Inland North, respectively. The recordings were completed in the years 2006–2008. None of the speakers reported any speech disorders. All participants were paid for their efforts.

B. Speech materials

Five American English vowels were selected for the study: /ɪ, ɛ, e, æ, aɪ/, which varied in their pattern of spectral change (or degree of diphthongization). Each vowel was contained in a target word of the structure /bVts/ and /bVdz/, which yielded the following words: *bits, bets, baits, bats, bites* and *bids, beds, bades, bads, bides*. The target word occurred in a sentence constructed to elicit two levels of vowel emphasis (emphatic and nonemphatic). In these sentences, only the main sentence stress varied, and the position of the target word and its immediate phonetic context remained unchanged. Thus, the proximity of the target word to the changing main sentence stress created the difference in the emphasis of the target word, as exemplified below:

Emphatic

Sue thinks the small CUTS are deep. No! Sue thinks the small **BITES** are deep.

Nonemphatic

Sue thinks the small **bites** are WIDE. No! Sue thinks the small **bites** are DEEP.

The use of sentence pairs rather than single sentences ensured fluency in reading, which was essential to examine the variation in the amount of the spectral change in vowels. It has been noted in preliminary studies that some speakers achieved the desired fluency while reading the second sen-

tence. Their first sentence tended to contain hesitations and pauses, which introduced noise to the clarity of exposition of the levels of emphasis. For this reason, only the second sentence in the pair was selected for the present analysis for a total of 1920 sentences (5 vowels × 2 consonantal contexts × 2 emphatic positions × 2 repetitions × 48 speakers).

C. Procedure

The testing took place at the university facilities in three locations (North Carolina, Ohio, and Wisconsin). A head-mounted Shure SM10A dynamic microphone was used, positioned at a distance of about 1.5 in. from the speaker's lips. The speaker was seated and was facing a computer monitor. Recordings were controlled by a custom program in MATLAB, which displayed the sentence pair to be read by the speaker and a set of control buttons for the experimenter. The sentences were presented in random order, and the recordings took place in one testing session. Speech samples were recorded and digitized at a 44.1-kHz sampling rate directly onto a hard disk drive. The speaker read the sentence pair placing the main sentence stress on the word in all caps. There was a short practice set completed before the start of the experiment. After recording each sentence pair, the experimenter either accepted and saved the utterance (which occurred most of the time) or re-recorded it in the case of any mispronunciations, disfluencies, or inaccurate stress placement. If the latter took place, the speaker was asked to repeat the utterance as many times as needed.

D. Acoustic measurements

The set of measurements included vowel duration and the frequencies of F1 and F2 over the course of vowel's duration, which were used to derive further measures of formant movement: VL, TL, and spectral roc. Prior to acoustic analysis, the tokens were digitally filtered and downsampled to 11.025 kHz.

1. Formant frequencies

Measurements of vowel duration served as input for subsequent automated measurements of formant frequencies at five equidistant temporal locations corresponding to the 20%–35%–50%–65%–80% point in the vowel. This was done to eliminate the immediate effects of surrounding consonants on vowel transitions and examine the variation in formant movement spanning over the vowel target. While proportional sampling of formants at two locations close to vowel onset and offset (i.e., 20%–80% or 20%–70%) or three locations including the temporal vowel midpoint (the 50% point) has been used more commonly in several acoustic studies (e.g., Ferguson and Kewley-Port, 2002; Hillenbrand *et al.*, 1995; Hillenbrand *et al.*, 2001), a denser multiple sampling at 4 (Fox, 1983), 9 (Adank *et al.*, 2004), or 16 equidistant points (Van Son and Pols, 1992) has also been done to estimate vowel inherent spectral change. The present use of five equidistant temporal points seeks to characterize the spectral change independent of vowel duration and provide enough information about formant trajectory changes which may be dialect-specific and may remain unnoticed while sampling the formants at only two or three points.

The frequency change in F1 and F2 over time was measured by centering a 25-ms Hanning window at each temporal location. F1 and F2 values were based on 14-pole linear predictive coding (LPC) analysis and were extracted automatically using a MATLAB program which displayed these values along with the fast fourier transform (FFT) and LPC spectrum and a wideband spectrogram of the entire vowel. In some cases, the formant values were verified using smoothed FFT spectra and wideband spectrograms with formant tracks displayed (using the program TF32, Milenkovic, 2003). Errors in formant estimation in LPC analysis were then hand-corrected.

2. Vowel duration

Standard measures of vowel duration were used (Peterson and Lehiste, 1960; Hillenbrand *et al.*, 1995). Vowel onsets and offsets were located by hand, primarily on the basis of a waveform display with segmentation decisions checked against a spectrogram. Vowel onset was measured from onset of periodicity (at a zero crossing) following the release burst of the stop (if present). In cases where closure remained voiced throughout and there was no evidence of an audible burst release, vowel onset was located at the point which indicated higher amplitude and higher frequency components. Vowel offset for words ending in a voiceless /ts/ was located at the point at which the amplitude of the vowel dropped to near zero (which was also coincident with elimination of all periodicity in the waveform). The vowel offset for words ending in a voiced /dz/ was defined as that point when the amplitude dropped significantly (to near zero). Since any voicing produced during the closure of a voiced stop will have relatively little high frequency energy (Pickett, 1999), this lack of high frequency components will produce a waveform that is relatively sinusoidal showing only slow variations (Olive *et al.*, 1993). When examining the waveforms, both cues were used to identify the location of the stop closure for /d/. All segmentation decisions were later checked and corrected (and then re-checked by a second experimenter) using a custom MATLAB program which displayed the segmentation marks superimposed over a display of the waveform (in two different views: a view that included the entire token and an expanded view that concentrated on the vowel portion only).

3. VL

VL, the length of a vector in F1 by F2 plane, is an indication of the amount of formant change in the course of vowel's duration, typically measured between the 20% and 80% points (Ferguson and Kewley-Port, 2002; Hillenbrand *et al.*, 1995). The assumption is that the longer the vector, the greater the magnitude of formant movement. Diphthongal or diphthongized vowels will have longer vectors than will monophthongs, which corresponds to their greater amount of frequency change. VL is included in the present study to assess its effectiveness as a measure of formant dynamics particularly for vowels in which the direction of formant

movement changes over time. VL is defined as a Euclidean distance (in hertz) between the 20% and 80% temporal points in the vowel in the F1 by F2 plane and is calculated as

$$\mathrm{VL} = \sqrt{(\mathrm{F1}_1 - \mathrm{F1}_5)^2 + (\mathrm{F2}_1 - \mathrm{F2}_5)^2}. \quad (1)$$

4. TL

Formant TL represents a measure of formant movement which tracks more closely formant frequency change over the course of vowel's duration than the magnitude of formant movement (VL). TL is potentially advantageous to measure the amount of frequency change for diphthongized and vowels whose curved formant tracks resemble a "U-turn" so that the values in the later portion of the vowel return to the values at the vowel's onset. Sampling formant frequencies at five equidistant locations allowed us to calculate TL for each of four separate vowel sections, i.e., 20%–35%, 35%–50%, 50%–65%, and 65%–80%, where the length of one vowel section (VSL) is

$$\mathrm{VSL}_n = \sqrt{(\mathrm{F1}_n - \mathrm{F1}_{n+1})^2 + (\mathrm{F2}_n - \mathrm{F2}_{n+1})^2}. \quad (2)$$

The overall formant TL was then defined as a sum of trajectories of four vowel sections:

$$\mathrm{TL} = \sum_{n=1}^{4} \mathrm{VSL}_n. \quad (3)$$

5. Spectral roc

Although TL measure can incorporate the curves in the formant tracks providing a detailed account of formant change, it fails to characterize the amount of frequency change over time. Yet, differences in vowel dynamics are manifested in the way the spectral change varies across vowel's duration. To address this, we first calculated the spectral roc (TL_roc) over the 60% portion of the vowel which was defined as

$$\mathrm{TL\_roc} = \frac{\mathrm{TL}}{0.60 \times \mathrm{v\_dur}}. \quad (4)$$

In addition, vowel section roc (VSL_roc) was calculated for each individual vowel section [determined by the temporal location of the five measurement points (20%–35%, 35%–50%, 50%–65%, and 65%–80%)] to compare regions of specific vowels and characterize the nature of the change within a particular region:

$$\mathrm{VSL\_roc}_n = \frac{\mathrm{VSL}_n}{0.15 \times \mathrm{v\_dur}}. \quad (5)$$

It is expected that VSL_roc will vary not only from section to section within a particular vowel but will also reveal potential differences in the way dialects utilize vowel dynamics for the same vowel "category."

TABLE I. Mean durations of individual vowels (in ms) (s.d.) in emphatic position preceding voiceless (b_vl) and voiced (b_vd) consonants.

Vowel	North Carolina b_vl	North Carolina b_vd	Ohio b_vl	Ohio b_vd	Wisconsin b_vl	Wisconsin b_vd
/ɪ/	170 (46)	226 (51)	125 (34)	185 (58)	106 (23)	150 (28)
/ɛ/	197 (45)	254 (57)	153 (38)	216 (60)	137 (23)	181 (32)
/e/	210 (49)	268 (57)	183 (36)	263 (63)	174 (29)	252 (44)
/æ/	251 (52)	292 (59)	229 (46)	300 (65)	215 (35)	277 (50)
/aɪ/	239 (39)	295 (52)	197 (38)	291 (68)	175 (26)	274 (52)
Total	214 (55)	267 (61)	178 (52)	251 (77)	162 (46)	227 (67)

III. RESULTS

A. Vowel duration

We begin with the presentation of the results for vowel duration. As displayed in Tables I and II, there were systematic differences in duration as a function of vowel quality, consonantal context, and degree of emphasis. Duration increased progressively with vowel openness which is a well-known intrinsic property of vowels. As also expected, vowels preceding voiced consonants were longer than before voiceless and emphatic vowels were longer than nonemphatic vowels. Of particular interest are differences in vowel duration as a function of dialect. North Carolina speakers produced the longest vowels, followed by Ohio and Wisconsin, respectively.

An analysis of variance (ANOVA) with the within-subject factor vowel, consonantal context and emphasis, and the between-subject factor dialect was used to assess these differences. For all reported significant main effects and interactions, the degrees of freedom for the F-tests were Greenhouse–Geisser adjusted in those cases in which there were significant violations of sphericity. In addition to the significance values, a measure of the effect size—partial eta squared (η^2)—is also reported.

All three within-subject effects were significant and their effect size was strong. As expected, the significant main effect of vowel ($[F(4,180)=674.37, p<0.001, \eta^2=0.937]$) reflected the intrinsic differences in the durations of the vowels examined here. The significant effect of consonantal context ($[F(1,45)=326.6, p<0.001, \eta^2=0.879]$) confirmed once again that vowel preceding a voiced consonant is longer than vowel preceding a voiceless consonant (means=213 and 166 ms, respectively). The significant effect of emphasis was

TABLE II. Mean durations of individual vowels (in ms) (s.d.) in nonemphatic position preceding voiceless (b_vl) and voiced (b_vd) consonants.

Vowel	North Carolina b_vl	North Carolina b_vd	Ohio b_vl	Ohio b_vd	Wisconsin b_vl	Wisconsin b_vd
/ɪ/	135 (33)	158 (41)	91 (24)	127 (40)	88 (18)	114 (37)
/ɛ/	153 (40)	166 (41)	117 (25)	130 (39)	113 (25)	121 (30)
/e/	178 (34)	206 (49)	148 (35)	185 (50)	144 (30)	175 (35)
/æ/	179 (38)	227 (56)	165 (36)	201 (46)	158 (26)	200 (36)
/aɪ/	194 (38)	225 (46)	164 (25)	210 (48)	156 (25)	206 (46)
Total	168 (42)	196 (55)	137 (41)	171 (56)	132 (37)	163 (53)

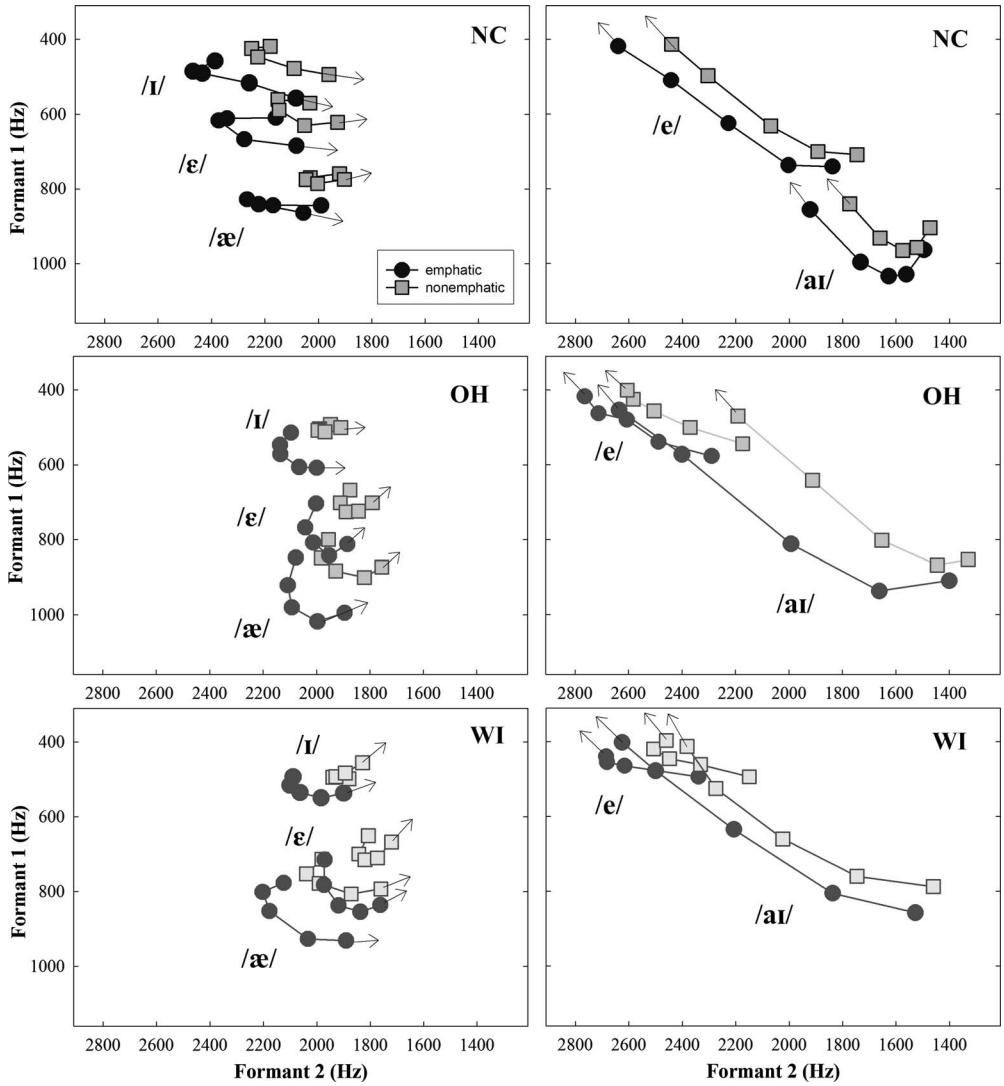


FIG. 1. Mean relative positions of monophthongal (left) and diphthongal (right) vowels and their formant movement measured at five equidistant points in the central 60% portion of the vowel. Shown are emphatic and nonemphatic vowels produced in the bVts context ("voiceless") by female speakers of three dialectal varieties of American English (spoken in Western North Carolina, Central Ohio, and Southern Wisconsin).

manifested in longer durations of emphatic vowels as compared to nonemphatic ([$F(1,45)=177.26$, $p<0.001$, $\eta^2=0.798$], means=217 and 161 ms, respectively). Interestingly, mean differences in vowel duration as a function of either emphasis or consonantal context were comparable (56 and 47 ms, respectively), indicating that consonantal context effects on vowel duration can be as great as the effects of emphasis.

The main effect of dialect was significant ([$F(2,45)=6.06$, $p=0.005$, $\eta^2=0.213$]) although its effect size was smaller than that for the within-subject factors. Subsequent *post-hoc* analyses using separate ANOVAs which included two dialects only showed that Wisconsin and North Carolina vowels differed significantly from one another (means=171 and 211 ms, respectively). However, Ohio vowels (means =185 ms) did not differ significantly from either Wisconsin or North Carolina vowels. These cross-dialectal differences in vowel duration are consistent with the results reported in Jacewicz *et al.* (2007) for young adults, confirming that dialectal differences in vowel duration do exist (at least for selected regions) and are independent of speaker age.

Several interactions were significant although their nature and small effect size do not warrant a separate discussion. One significant interaction between context and emphasis deserves mention given its large effect size [$F(1,45)=152.29$, $p<0.001$, $\eta^2=0.772$]. The interaction arose from the fact that emphatic vowels in the context of voiced consonants were substantially longer (72 ms or 41%) than nonemphatic vowels in this environment whereas the emphasis-related difference for vowels preceding voiceless consonants was smaller (39 ms or 27%).

B. Formant movement

Turning to formant analysis, Figs. 1 and 2 display relative positions in the F1×F2 plane and formant movement of vowels preceding voiceless and voiced consonants, respectively. The left panels show "monophthongal" vowels /ɪ, ɛ, æ/ and the right panels the "diphthongal" /e, aɪ/. Direction of formant movement is indicated by arrows.

Based on visual inspection, there is a substantial variation in formant dynamics across individual vowels and dia-

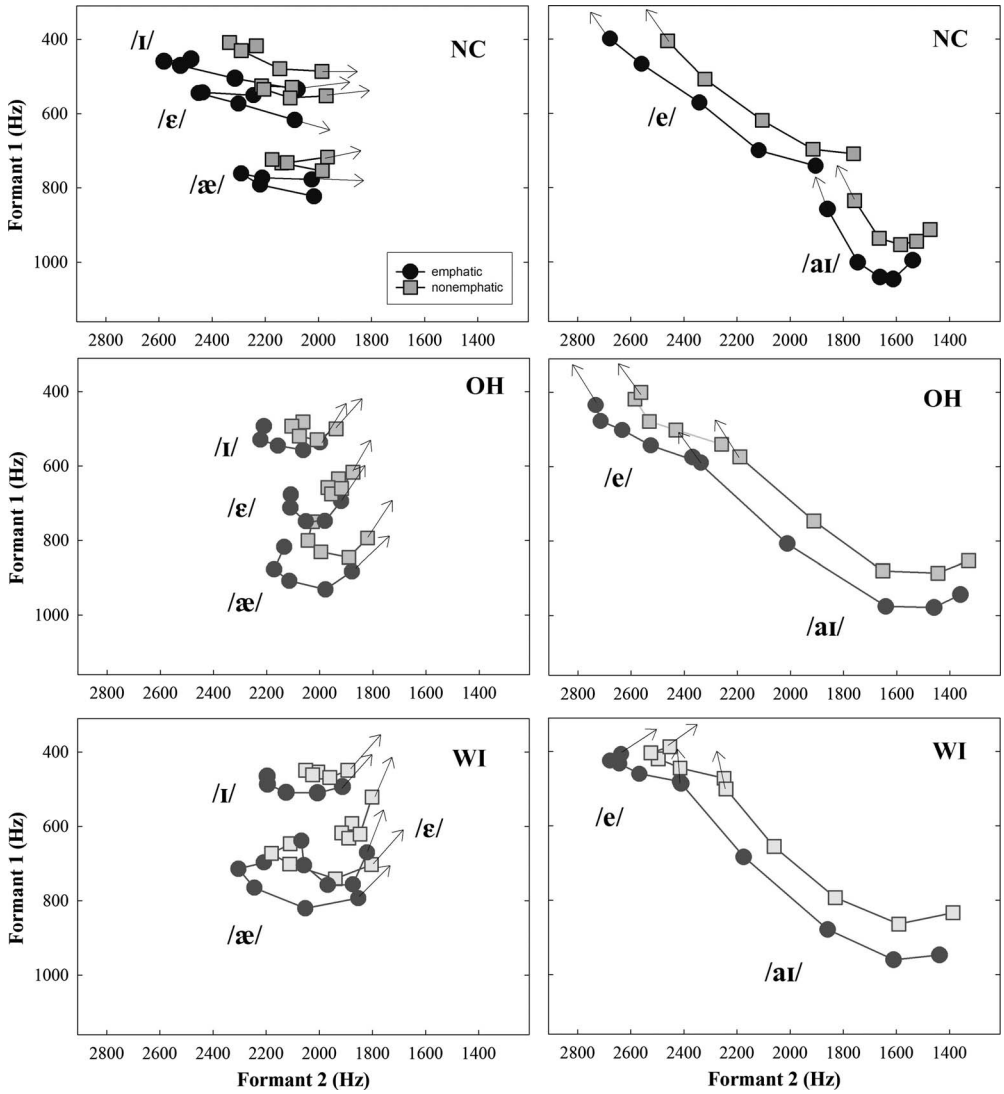


FIG. 2. Mean values of the vowels in the bVdz ("voiced") context (see Fig. 1 legend for details).

lects. North Carolina /ɪ, ɛ, æ/ are the most fronted with the nature of their formant movement distinct from both Ohio and Wisconsin vowels. Across all dialects, emphatic vowels are more peripheral and show more formant movement than nonemphatic vowels. Cross-dialectal differences are particularly evident for /e, aɪ/. The North Carolina /e/ is the most diphthongal and the Wisconsin /e/ may even be regarded as a monophthong given its small amount of change in F1. The diphthong /aɪ/, on the other hand, is relatively monophthongal in North Carolina but shows a great amount of spectral change in both Ohio and Wisconsin. The Wisconsin /æ/ is raised due to the Northern Cities Shift and it can be seen that its nonemphatic variant has considerable overlap with the emphatic /ɛ/.[1]

It needs to be underscored that the formant track dynamics displayed in Figs. 1 and 2 is plotted from measurements at five temporally equidistant points during a vowel. Thus, these frequency measurements are time-normalized across all vowels and do not reflect differences in vowel duration. This issue will be addressed subsequently.

C. VL

The first measure applied to assess the present variation in formant dynamics is VL (e.g., Ferguson and Kewley-Port, 2002; Hillenbrand *et al.*, 1995; Hillenbrand and Nearey, 1999). As Fig. 3 shows, VLs are smaller for some vowels such as /ɪ, ɛ/ but every vowel exhibits at least some amount of spectral change. There are clear VL differences as a function of dialect, especially between North Carolina vowels and those from the two Midwestern dialects. A separate repeated-measures ANOVA was conducted for each vowel[2] with the within-subject factors consonantal context and emphasis. Dialect was included as the between-subject factor. In general, all three main effects were significant. One exception was the vowel /ɛ/, whose VLs did not differ significantly as a function of dialect. Table III summarizes the results of the analyses. As can be seen, the effect size was typically greater for the main effect of emphasis compared to the main effect of consonantal context. For all vowels, VLs were significantly longer for emphatic vowels than for nonemphatic. The context effects were more variable, indicating longer

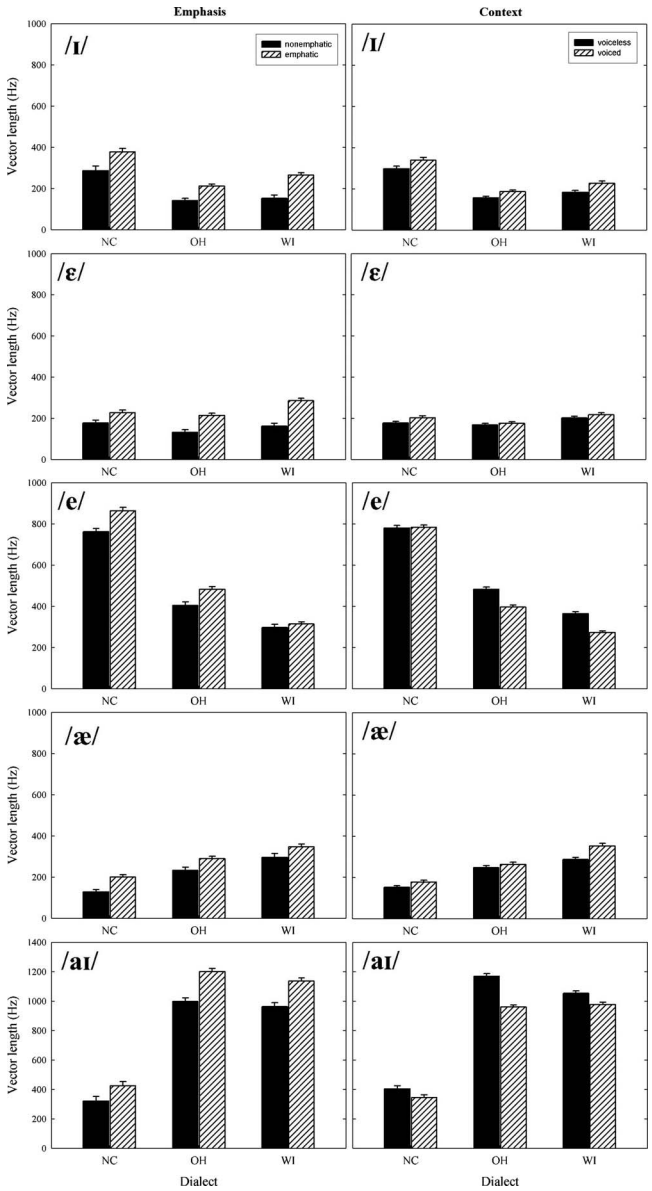


FIG. 3. Mean values (s.e.) for VL, i.e., F1 and F2 frequencies change between the 20%–80% temporal point for each vowel in each dialect as a function of vowel emphasis and consonantal context.

VLs when the vowel was followed by voiced consonants in the case of /ɪ, ɛ, æ/ and longer VLs when it was followed by voiceless consonants for /e, aɪ/. The effects of dialect were particularly strong for the vowels /e, aɪ/ due to the fact that North Carolina VLs were clearly different from both Midwestern variants. The North Carolina /e/ had the longest VL which was more than twice that of the Wisconsin variant, with Ohio vowel falling in between. For the diphthong /aɪ/, Ohio variant had the longest VL, about three times that of North Carolina /aɪ/ which is well known for being relatively monophthongal in this regional variety of English.

It can be expected that the large differences between the North Carolina vowels and the vowels from the two Midwestern varieties will produce a significant main effect of dialect. However, the differences between the Ohio and Wisconsin vowels themselves may be too small to reach significance. Additional repeated-measures ANOVAs were used to examine the significance of all three factors (emphasis, context, and dialect) for Ohio and Wisconsin vowels while excluding North Carolina from the analyses. As expected, the results showed significant main effects of emphasis and context for each of the five vowels. However, the main effect of dialect was significant only for /e/ $[F(1,30)=19.55,\ p<0.001,\ \eta^2=0.394]$, indicating longer VL for Ohio variant compared to Wisconsin.

In summary, the present results show that VLs varied significantly with vowel emphasis and consonantal context, and dialectal differences were also apparent, at least between North Carolina and Midwestern vowels. However, one issue needs to be resolved before accepting VL as a measure which characterizes the true amount of frequency change in the course of vowel's duration. In particular, one can argue that VL, in fact, underestimates the amount of spectral change in a vowel and may lead to false interpretations of the nature of the spectral change being examined. To exemplify the point, we will now consider two examples of North Carolina vowels: /æ/ and /e/.

The left panel of Fig. 4 shows the North Carolina variant of /æ/ redrawn here from Fig. 1 for the purposes of illustration. VL is a measure of formant frequency change between the 20% and 80% points in the vowel. As evident, VL fails to

TABLE III. Summary of significant main effects and interactions from repeated measures ANOVAs for VL. Shown are partial eta squared values (η^2).—not significant, vd=voiced, vl=voiceless, e=emphatic, ne=nonemphatic, NC=North Carolina, OH=Ohio, and WI=Wisconsin.

	/ɪ/	/ɛ/	/e/	/æ/	/aɪ/
Context	0.269[a]	0.152[b]	0.515[a]	0.284[a]	0.267[a]
	vd>vl	vd>vl	vl>vd	vd>vl	vl>vd
Emphasis	0.474[a]	0.517[a]	0.375[a]	0.325[a]	0.510[a]
	e>ne	e>ne	e>ne	e>ne	e>ne
Dialect	0.305[a]	—	0.818[a]	0.292[a]	0.800[a]
	NC>WI>OH		NC>OH>WI	WI>OH>NC	OH>WI>NC
Context×Emphasis	0.092[c]	—	—	—	0.111[c]
Context×Dialect	—	—	0.341[a]	—	—
Emphasis×Dialect	—	—	0.155[c]	—	—
Con×Emp×Dialect	—	—	0.136[c]	—	—

[a] $p<0.001$.
[b] $p<0.010$.
[c] $p<0.050$.

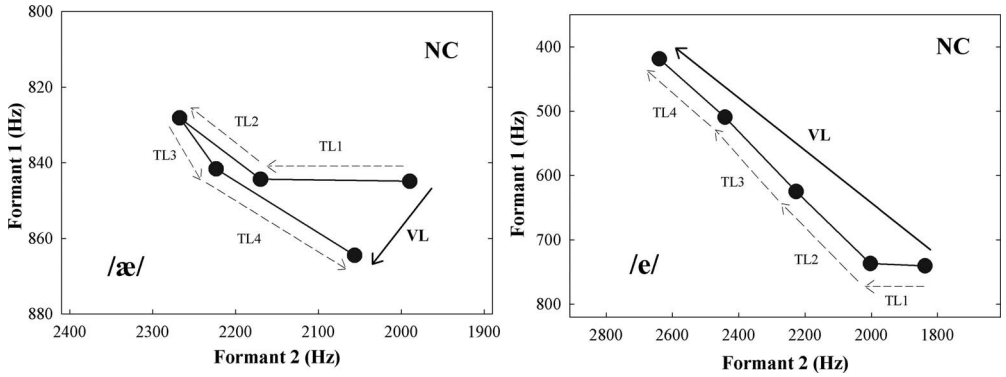


FIG. 4. Measurement of VL and total TL for North Carolina variant of /æ/ (left) and /e/ (right) in emphatic positions redrawn from Fig. 1.

account for the actual formant movement over time. The length of the entire formant trajectory consists here of four sections (TL1–TL4), each corresponding to formant change between two consecutive measurement points (20%–35%, 35%–50%, 50%–65%, and 65%–80%). Because the trajectory of North Carolina /æ/ is U-shaped (which reflects the "Southern drawl"), the VL estimate is particularly inadequate to measure this type of spectral change. Yet, VL can be quite accurate in assessing diphthongal changes such as for North Carolina /e/ shown in the right panel. This vowel, also redrawn from Fig. 1, shows an almost linear spectral change across its four sections. Thus, the estimated total trajectory change can be approximated relatively well by the length of the vector which expresses a linear distance between the 20%- and 80% points.

In summary, VL does appear to capture some aspects of formant movement and is rather reliable as a measure of linear trajectory change. However, formant trajectory shapes can vary cross-dialectally in ways impossible to characterize by the use of VL. The North Carolina /æ/ is the most fitting example. It seems that computing the length of the entire trajectory, i.e., approximated by the multiple-point sampling, may account more reliably for the extent of spectral change in a vowel. Section IV will address this possibility.

D. TL

The total TL, consisting of the sum of TLs of the four vowel sections, is expected to provide a more detailed estimate of formant change. Figure 5 shows mean TL values for each vowel broken down by emphasis and consonantal context. As expected, the TL values are greater than those for VL in Fig. 3. As it was done for the VL measure, a repeated-measures ANOVA with the within-subject factors emphasis and consonantal context and between-subject factor dialect was conducted for each vowel.

The main effects of emphasis and consonantal contexts were significant, and the general results were in accord with those for VL: emphatic vowels had significantly greater TLs than nonemphatic vowels, the vowels /ɪ, ɛ, æ/ had longer TLs when followed by voiced consonants, and /e, aɪ/ had longer TLs when followed by voiceless consonant. Table IV summarizes the results of the analyses.

The effects of dialect were somewhat different for TLs, however. Although the main effect of dialect was significant for the vowels /ɪ, e, aɪ/ and the order of dialectal variants in terms of the amount of the spectral change were in agreement with the results for VL (including the significant difference between the Ohio and Wisconsin /e/), discrepancies between the two measures were found for the vowels /ɛ/ and

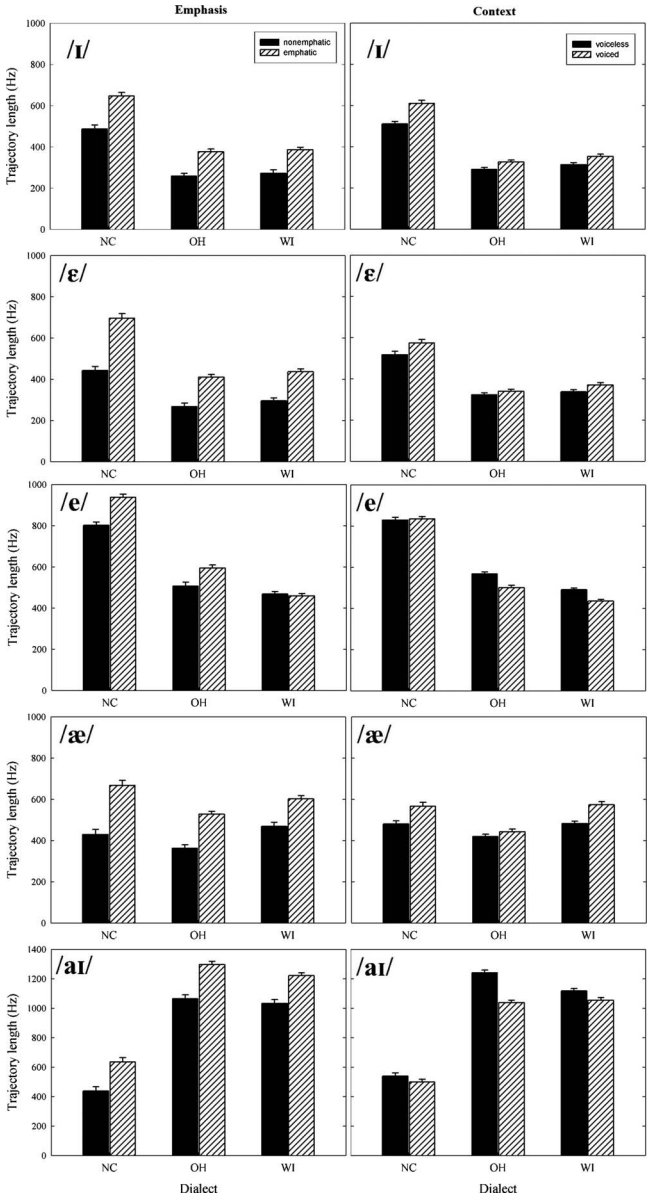


FIG. 5. TL, i.e., sum of the VSLs of four vowel sections over the central 60% of the vowel. Shown are mean values (s.e.) for each vowel in each dialect as a function of vowel emphasis and consonantal context.

TABLE IV. Summary of significant main effects and interactions from repeated measures ANOVAs for total TL. Shown are partial eta squared values (η^2).—not significant, vd=voiced, vl=voiceless, e=emphatic, ne=nonemphatic, NC=North Carolina, OH=Ohio, and WI=Wisconsin.

	/ɪ/	/ɛ/	/e/	/æ/	/aɪ/
Emphasis	0.581[a]	0.651[a]	0.319[a]	0.680[a]	0.623[a]
	e > ne	e > ne	e > ne	e > ne	e > ne
Context	0.503[a]	0.118[b]	0.266[a]	0.402[a]	0.280[a]
	vd > vl	vd > vl	vl > vd	vd > vl	vl > vd
Dialect	0.670[a]	0.516[a]	0.804[a]	—	0.737[a]
	NC > WI > OH	NC > WI > OH	NC > OH > WI	—	OH > WI > NC
Context × Emphasis	—	—	—	—	—
Context × Dialect	0.138[b]	—	0.180	—	0.180[b]
Emphasis × Dialect	—	0.144[b]	0.257[c]	—	—
Cont × Emp × Dialec	0.139[b]	—	—	—	—

[a] $p < 0.001$.
[b] $p < 0.050$.
[c] $p < 0.010$.

/æ/. In particular, there was no significant effect of dialect for the VL measure for /ɛ/ [$F(2,45)=1.56$, $p=0.221$, $\eta^2=0.065$], whereas dialect was significant for TL [$F(2,45)=23.95$, $p<0.001$, $\eta^2=0.516$], showing greater TLs for North Carolina /ɛ/ (mean=571 Hz) than for either Wisconsin (mean=366 Hz) or Ohio (mean=341 Hz). For the vowel /æ/, the pattern was reversed: the main effect of dialect was significant for the VL measure [$F(2,45)=9.29$, $p<0.001$, $\eta^2=0.292$], showing greatest VLs for Wisconsin (mean=317 Hz) followed by Ohio and North Carolina (means=259 and 164 Hz, respectively). For the TL measure, dialect was not significant [$F(2,45)=3.15$, $p=0.053$, $\eta^2=0.123$] and North Carolina /æ/ had slightly greater TL (mean=549 Hz) than Wisconsin (mean=535 Hz), with Ohio falling last (444 Hz). Clearly, these discrepancies arose from underestimating the amount of formant change by the VL measure due to the change in the direction of formant curves.

To compare the results of the two measures, i.e., VL and TL, separate repeated-measures ANOVAs were used for each vowel and for each dialect with the within-subject factors formant change (VL, TL), emphasis, and consonantal context. Table V summarizes the results for the main effect of formant change.

As can be seen, the differences between VL and TL were highly significant for each vowel in each dialect. Next to the effect size, the table lists in parentheses the percentage of underestimation of formant change by the VL measure. The underestimation was found to be as great as 70% for the North Carolina /æ/ and as small as 7%–8% for the Ohio and Wisconsin /aɪ/ and North Carolina /e/. For the remaining vowels, the VL underestimation ranged from 19% to 65%. These results show an advantage of the TL measure over VL, especially for vowels which exhibit a change in the direction of formant movement. The general picture of TL advantage for each vowel averaged across emphasis levels and consonantal contexts can be found in Fig. 6. For each dialect, the VL underestimation of formant change for the vowels /ɪ, ɛ, æ/ is considerably greater than for the diphthongal vowels /e, aɪ/. These differences were found for each dialect, indicating that the TL measure reflects dialect-specific spectral change in vowels quite well.

In summary, the statistical evidence along with the graphic displays suggests that VL does not account reliably for the dialectal differences. The extent of formant movement is better characterized by a TL measure, which utilizes formant measurements sampled at multiple points in a vowel.

E. Spectral roc

Although the TL measure appears to be more reliable in addressing dialectal differences, the measurement points are time normalized and indicate only relative positions across the vowel. This, of course, fails to account for how quickly (or slowly) these formant frequency changes occur in time. Yet, there may be important dynamic differences across dialects, contexts, and speaker age that relate to such spectro-temporal changes. The spectral roc measure presented here will allow us to make these comparisons.

TABLE V. Summary of the significant main effect of formant change (VL vs TL) from repeated measures ANOVAs. Shown are partial eta squared values (η^2), The values in parentheses indicate percentage of underestimation of formant movement by the VL measure.

	/ɪ/	/ɛ/	/e/	/æ/	/aɪ/
North Carolina	0.895[a] (42)	0.893[a] (65)	0.792[a] (7)	0.876[a] (70)	0.855[a] (31)
Ohio	0.907[a] (43)	0.898[a] (49)	0.744[a] (19)	0.917[a] (42)	0.896[a] (7)
Wisconsin	0.910[a] (36)	0.916[a] (39)	0.814[a] (34)	0.945[a] (41)	0.878[a] (8)

[a] $p < 0.001$.

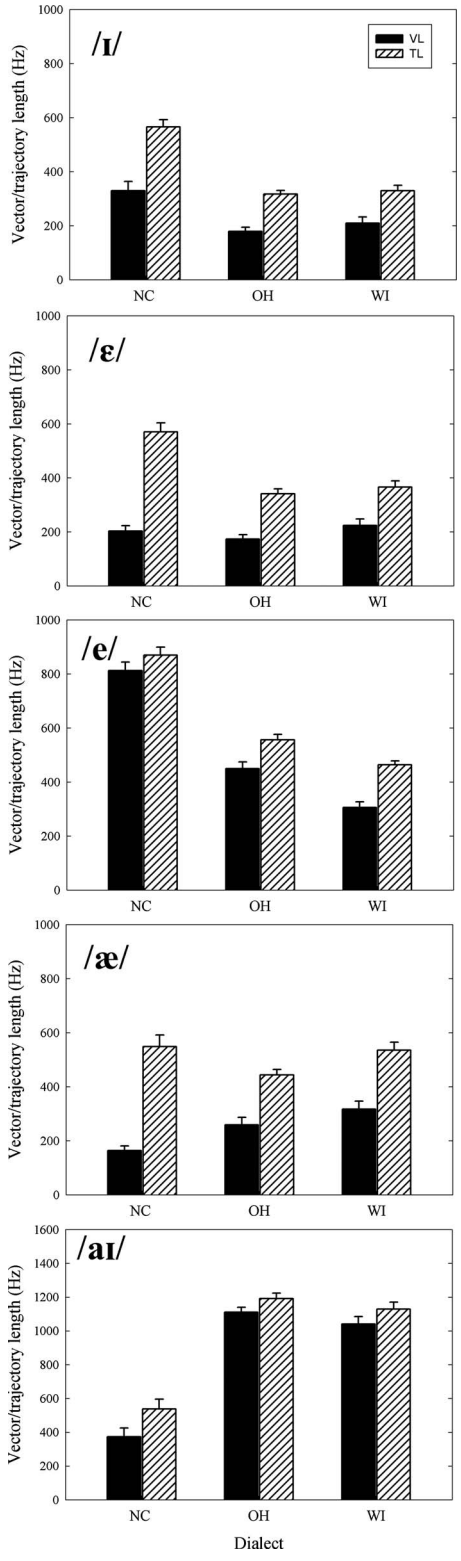


FIG. 6. A comparison of VL and TL for each vowel and each dialect. Shown are mean values (s.e.) averaged across emphasis levels and consonantal context.

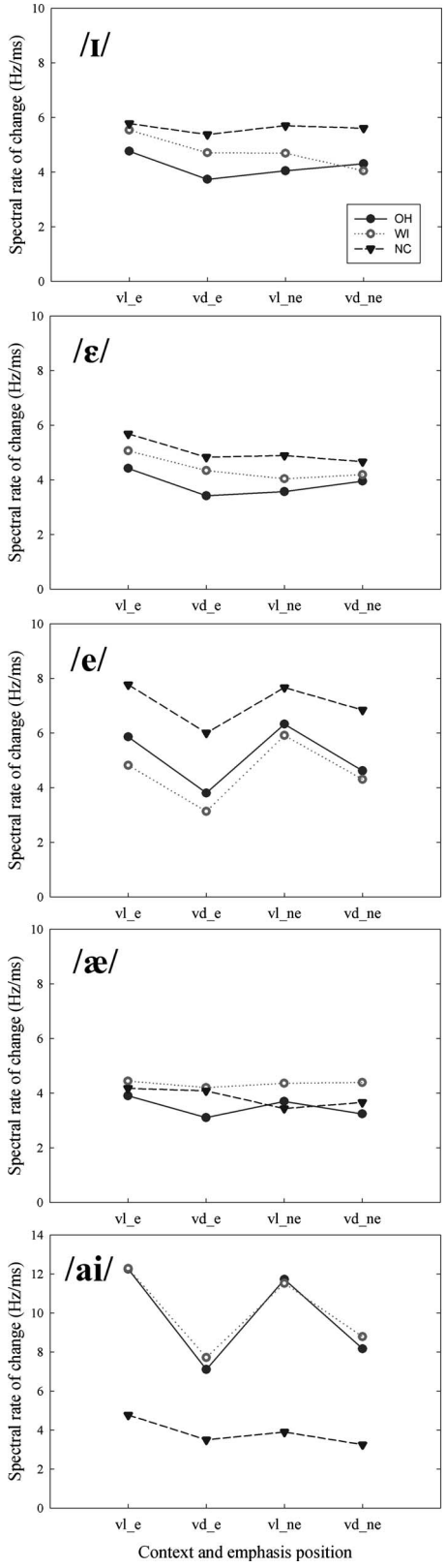


FIG. 7. Mean spectral roc at the targets of vowels in variable emphasis positions (e=emphatic, ne=nonemphetic) and consonantal contexts (vl =voiceless, vd=voiced) across the dialects.

Shown in Fig. 7 is the spectral roc for the five vowels in both emphatic and nonemphatic positions in voiced and voiceless contexts for each of the three dialects. As might be expected, overall spectral roc varies as a function of vowel category. The mean values were highest for the diphthong /aɪ/ in both Wisconsin (10.1 Hz/ms) and Ohio (9.8 Hz/ms) varieties (but not in North Carolina, 3.9 Hz/ms) and lowest for the vowel /æ/ in each of the three dialects (4.4, 3.5, and 3.8 Hz/ms, respectively). Dialectal differences were particularly evident in the case of /e/ which had the highest spectral

TABLE VI. Summary of significant main effects and interactions from repeated measures ANOVAs for spectral roc. Shown are partial eta squared values (η^2).—not significant, vd=voiced, vl=voiceless, e=emphatic, ne=nonemphatic, NC=North Carolina, OH=Ohio, and WI=Wisconsin.

	/ɪ/	/ɛ/	/e/	/æ/	/aɪ/
Emphasis	—	0.094[a]	0.262[b]	—	—
		e>ne	ne>e		
Context	0.163[c]	0.160[c]	0.754[b]	0.089[a]	0.783[b]
	vl>vd	vl>vd	vl>vd	vl>vd	vl>vd
Dialect	0.286[c]	0.211[c]	0.484[b]	0.154[a]	0.745[b]
	NC>WI>OH	NC>WI>OH	NC>OH>WI	WI>NC>OH	WI>OH>NC
Context×Emphasis	0.134[a]	0.288[b]	0.106[a]	0.089[a]	0.353[b]
Context×Dialect	—	—	—	0.146[a]	0.465[b]
Emphasis×Dialect	—	—	—	—	—
Cont×Emp×Dialect	—	—	—	—	—

[a] $p<0.05$.
[b] $p<0.001$.
[c] $p<0.010$.

roc among all North Carolina vowels (mean=7.1 Hz/ms) and second highest among the Ohio vowels (mean =5.1 Hz/ms). In Wisconsin, however, the mean value was lower (4.5 Hz/ms) and it was comparable with roc for /ɛ/ and /æ/ (both 4.4 Hz/ms).

A separate repeated-measures ANOVA with the within-subject factors emphasis and consonantal context and the between-subject factor dialect was conducted for each vowel. As summarized in Table VI, vowel emphasis did not have a strong effect on spectral roc. Rather, it was the consonantal context that affected the spectral roc of all vowels in a systematic way: without exception, vowels preceding voiceless consonants had higher spectral roc than when preceding voiced consonants. This effect was particularly strong for the vowels /e, aɪ/. The main effect of dialect was significant for each vowel. For the vowels /ɪ, ɛ, e/ North Carolina had the highest spectral roc among the three dialects. For /æ, aɪ/, the spectral roc was highest in the variety of English spoken in Wisconsin. Also significant for each vowel was the interaction between context and emphasis. This interaction, although not particularly strong, was manifested somewhat differently for monophthongal /ɪ, ɛ, æ/ and diphthongal /e, aɪ/ vowels.[3]

Of particular interest to this study are dialectal differences in the spectral roc which persisted when two additional sources of contextual variation in roc were included, i.e., degree of vowel emphasis and the type of consonantal context. Clearly, changes in spectral roc arise from differences in vowel duration, TL or a combination of the two. In an attempt to better understand the contribution of each source of this variation and explain the obtained patterns, we will now examine proportional differences in vowel duration and TL which arose from the two contextual factors, vowel emphasis and consonantal context.

Listed in Table VII are changes in vowel duration and TL as a function of vowel emphasis. Of interest are percentages of reduction in vowel duration in nonemphatic positions and corresponding reduction in TL for each vowel in each dialect. The general tendency is that the proportion of reduction in duration of nonemphatic vowels corresponds roughly to the proportion of reduction in their TLs, which will not affect their spectral roc. This would explain the lack of significant effect of emphasis on the spectral roc, at least for three out of five vowels. A different outcome was found for the consonantal context effects, as summarized in Table VIII. Here the relative reduction in TL for vowels in voiceless context tends to be smaller than the reduction in vowel duration, which produces higher spectral roc in the voiceless context compared to the voiced context. These results are in accord with the number of significant effects of consonantal context on the spectral roc of individual vowels (compare Table VI).

A word of caution against an exclusive reliance on these general trends in explaining the spectral roc results is needed, however. Several factors interact here, and each has some impact on the movement of articulators which is the underlying source of variation in the spectral roc. In some cases, it is the dialect-specific spectral change that interacts differently with contextual factors. To illustrate the point, we will consider one example here, that of proportional reductions in both vowel duration and TL for the vowel /æ/, which have been found to vary across dialects.

As Table VIII indicates, the North Carolina variant does not increase its spectral roc in the voiceless context. Rather, spectral roc increases in emphatic position which has a greater TL (and smaller decrease in duration) as compared to the nonemphatic position (see Table VII). However, consonantal context (and not emphasis) affected spectral roc of the Ohio /æ/ which increased in the voiceless context due to the small reduction in TL. Finally, the proportional reductions as a function of both emphasis and context did not vary much for the Wisconsin variant, suggesting that, in this dialect, spectral roc of /æ/ does not change across emphatic positions and contexts. Results of separate ANOVAs used for each dialect support this explanation. The main effect of emphasis was significant for the North Carolina /æ/ ($[F(1,15)=4.97$, $p=0.041$, $\eta^2=0.249]$) and indicated higher spectral roc in emphatic position. The main effect of context was significant for Ohio /æ/ ($[F(1,15)=7.15$, $p=0.017$, $\eta^2=0.323]$), showing higher spectral roc in the voiceless context. Finally, nei-

TABLE VII. Changes in mean values of vowel duration (vow dur) and TL as a function of vowel emphasis along with percentages of their reductions in nonemphatic relative to emphatic positions.

Vowel	Vow dur emphatic (ms)	Vow dur nonemphatic (ms)	Vow dur reduction ne<e (%)	TL emphatic (Hz)	TL nonemphatic (Hz)	TL reduction ne<e (%)
			North Carolina			
/ɪ/	197.2	146.7	25.6	645.4	486.5	24.6
/ɛ/	225.5	159.4	29.3	699.3	442.0	36.8
/e/	239.0	191.7	19.8	936.8	801.5	14.4
/æ/	271.4	202.7	25.3	667.4	429.6	35.6
/aɪ/	267.1	209.2	21.7	634.7	442.3	30.3
			Ohio			
/ɪ/	157.1	109.2	30.5	376.8	257.7	31.6
/ɛ/	186.7	123.9	33.6	411.9	269.7	34.5
/e/	225.6	168.3	25.4	595.4	517.6	13.1
/æ/	266.4	183.6	31.1	521.9	365.8	29.9
/aɪ/	245.1	188.4	23.1	1308.9	1075.0	17.9
			Wisconsin			
/ɪ/	128.5	101.1	21.3	386.3	272.4	29.5
/ɛ/	158.6	117.2	26.1	436.6	294.6	32.5
/e/	213.0	159.4	25.2	460.2	468.0	−1.7
/æ/	246.3	178.9	27.4	602.2	468.4	22.2
/aɪ/	224.7	181.0	19.5	1220.5	1036.5	15.1

ther emphasis nor context was significant for the Wisconsin variant. No other effects and interactions were significant in these analyses.

These results support the claim that, as a measure, spectral roc is sensitive to dialectal differences in vowel dynamics and can provide details of complex interactions of several factors. Since roc does not require extensive computations, it can be used effectively in analyzing a larger corpus.

IV. DISCUSSION

The present study sought to characterize the nature of the dynamic spectral change found in the targets of selected

TABLE VIII. Changes in mean values of vowel duration (vow dur) and TL as a function of consonantal context along with percentages of their reductions in voiceless relative to voiced contexts.

Vowel	Vow dur voiced (ms)	Vow dur voiceless (ms)	Vow dur reduction vl<vd (%)	TL voiced (Hz)	TL voiceless (Hz)	TL reduction vl<vd (%)
			North Carolina			
/ɪ/	191.4	152.6	20.3	618.6	513.3	17.0
/ɛ/	209.8	175.0	16.6	585.7	555.6	5.1
/e/	236.9	193.9	18.1	871.0	867.3	0.4
/æ/	259.3	214.8	17.1	603.9	493.1	18.3
/aɪ/	259.8	216.6	16.6	524.2	552.8	−5.5
			Ohio			
/ɪ/	156.8	109.5	30.2	350.3	284.2	18.9
/ɛ/	173.5	137.1	21.0	354.9	326.7	7.9
/e/	225.1	168.7	25.0	526.5	586.5	−11.4
/æ/	251.4	198.5	21.1	459.1	428.5	6.7
/aɪ/	251.3	182.2	27.5	1086.0	1298.0	−19.5
			Wisconsin			
/ɪ/	132.2	97.4	26.3	348.2	310.5	10.8
/ɛ/	150.8	125.0	17.1	387.4	343.7	11.3
/e/	213.2	159.1	25.4	427.9	500.4	−17.0
/æ/	238.8	186.5	21.9	587.7	482.9	17.8
/aɪ/	239.8	165.8	30.8	1092.4	1164.5	−6.6

American English vowels in three distinct dialectal regions in the United States. The results are encouraging in that we are beginning to find ways to better understand vowel dynamics across different American English dialects. In particular, we found cross-dialectal differences in vowel duration, in the extent of spectral change in formant trajectories, and in the spectral roc.

Although our primary goal was to find a set of effective measures which would reveal systematic differences among the regional variants, the study also gained more insights into positional relations within the vowel system of each dialect. As seen in Figs. 1 and 2, Ohio and Wisconsin variants of /ɛ, æ/ tend to spectrally overlap when variable emphasis is taken into consideration: the emphatic /ɛ/ approximates the position of the nonemphatic /æ/. This is not the case for North Carolina, where /ɛ/ and /æ/ are clearly separated under variable emphasis conditions and a possibility of an overlap arises for /ɪ, ɛ/ rather. The nature of formant dynamics is also highly variable across the dialects, and the magnitude of formant movement can vary dramatically such as for the diphthongal vowels /e, aɪ/.

A. Characterizing the variation in formant dynamics

We first turned to an established procedure of estimating the magnitude of formant frequency change between the 20% and 80% temporal points in a vowel (VL). Although some of the spectral variation could be accounted for by this measure, we excluded it from further consideration as it did not provide a satisfying characterization of the most dynamic spectral changes in several of the vowels. In particular, the extent of formant movement was greatly underestimated for vowels in which the direction of this movement in the F1 by F2 space changes over time. We found that the total trajectory change (TL) over multiple temporal locations for the vowel center represents more adequately the magnitude of formant movement. Relating the spectral change over the total trajectory to the time necessary to execute this formant movement, we computed the spectral roc which provides another view of the time-varying information in a vowel.

Two phonetic factors that affect vowel duration, emphatic stress and the voicing status of the consonant that follows the vowel, were systematically varied in this study. Entered as within-subject factors, the two sources of variation were found to interact with formant movement in somewhat different way: while both the emphatic vowel and vowel preceding a voiced consonant had longer durations, the spectral roc was significantly higher for the shorter vowel followed by a voiceless consonant and not for the shorter nonemphatic vowel. The small effect size of emphasis found in the analyses of spectral roc will need to be investigated separately in greater detail. We can only speculate that this variation comes from the way emphatic stress is brought about by vowel-specific articulatory actions.[4]

Apart from the variation in vowel dynamics coming from phonetic sources, dialect was found to be a strong source of variation in vowel-inherent spectral change. The effects of dialect were found for each vowel examined in the present study. As an example, an interesting relationship between vowel duration and the dialect-specific nature of formant trajectory change was found for the vowel /ɪ/. North Carolina /ɪ/ was longer, had a greater TL, and faster spectral roc (means: 172 ms, 566 Hz, 5.6 Hz/ms) than Ohio /ɪ/ which was shorter, had a smaller TL, and slower spectral roc (means: 133 ms, 317 Hz, 4.2 Hz/ms). These differences stem from the nature of the dynamic formant changes in each dialect which, in very general terms, are brought about by faster articulatory gestures in order to produce the North Carolina vowel and comparatively slower gestures in a slightly diphthongal variety of the Ohio /ɪ/.

The spectral roc measure used in this study is just one possible measure which, to some extent, reflects speed of articulatory movement over the course of the vowel's target. Although this measure can only give an indirect indication of the speed of specific articulators underlying the production of diphthongal and quasi-diphthongal changes, it is nevertheless useful in estimating the average pace of formant movement during the central 60% of the total spectral change. A related measure of spectral change, although assessing F2 velocity only, was used in Moon and Lindblom (1994). A more detailed measurement such as by fitting linear regression lines to the formants and computing the slopes (Wouters and Macon, 2002a) will be problematic in this particular set of data because of the changing directionality of the formant movement. While Wouters and Macon (2002a) studied liquid-vowel and vowel-liquid transitions and diphthong transitions in the productions of one speaker, this approach will not be effective in dealing with the type of spectral changes such as those found in North Carolina vowels. The present approach, being relatively easy to implement, can be more readily used in a sociophonetic setting which, by definition, must involve a larger corpus of data. Having established the types of variation in formant trajectories that can be expected in cross-dialectal data in terms of TL, directionality, and curvature, a refinement of the current measures will be undertaken in order to address the changes in the direction of formant movement. In particular, parametrization procedures can be used (e.g., Harrington, 2006; Harrington *et al.*, 2008; Hillenbrand *et al.*, 2001; Morrison, 2009; Zahorian and Jagharghi, 1993) in order to model the various trajectory shapes.

It is clear that the nature and amount of spectral change for vowels studied here can be characterized quite effectively when formant trajectories are sampled at multiple time points rather than at one temporal location at the vowel target. The use of five temporal locations estimates the dynamic trajectory to the extent that time-normalized spectral variation can be assessed rather accurately. The addition of the time dimension and inclusion of the spectral roc provides further insights as to how vowel-inherent spectral change differs for individual vowels across several regional variants. Thus, the combination of the three basic acoustic parameters (TL, vowel duration, and spectral roc) can be effective in characterizing the regional variation in American English vowels.

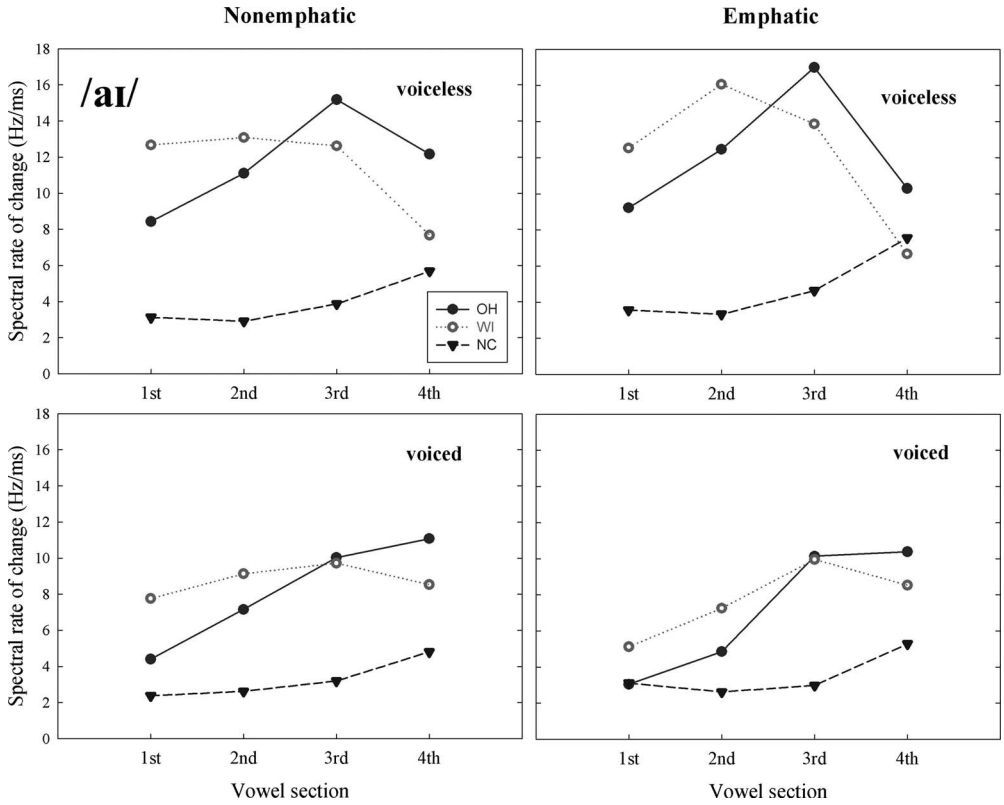


FIG. 8. Mean spectral roc for the diphthong /aɪ/ in the four consecutive sections of the target in nonemphatic (left) and emphatic (right) positions across the dialects.

B. Dialectal differences in the dynamics of /aɪ/

Regional dialect turns out to be a rich source of phonetic variation. The diphthong /aɪ/ is a good example of how differently vowels can be produced from region to region. This diphthong can be almost monophthongal in North Carolina and it can have two manifestations in even closely related Midwestern dialects spoken in Ohio and Wisconsin. It is interesting to note how the proportional differences in duration and TL arising from the effect of consonantal context interact with dialect-specific variation in spectral roc.

As shown in Table VI, the effects of consonantal context and dialect on the spectral roc of /aɪ/ were particularly strong. From Table VIII we find that while vowel duration was reduced in voiceless context (especially for the Ohio and Wisconsin variants), the TL increased greatly. The shorter duration and greater TL affected the spectral roc which was actually higher in the voiceless than the voiced context (see the negative percentage values). This apparent divergence from the expected pattern of reduction in the voiceless context supports earlier reports in the literature, however. For example, Gay (1968) found that the increased duration of /aɪ/ in *bide* relative to *bite* is accomplished primarily by a lengthening of the steady-state onset of the diphthong while both the gliding portion and the diphthongal offset lengthen to a lesser extent. The lengthening of the onset was verified by Jacewicz *et al.* (2003). This study also showed that the F2 change in bide was smaller mostly due to lower terminal frequency values of the offglide. A larger F2 change coupled with a shorter duration means that *bite* will have a higher spectral roc than *bide*. The present results are in accord with these previous findings and can be easily inferred from the patterns in /aɪ/ displayed in Figs. 1 and 2. That is, there is a smaller spectral change in the first two sections (i.e., over the first three data points) of the Ohio and Wisconsin diphthongs in the voiced context and a comparatively greater formant movement in the voiceless context. Given shorter vowel duration in the latter context, we can thus explain the much higher spectral roc in the voiceless context than in the voiced.

There were also notable dialectal differences in the production of the diphthong. Figure 8 gives a more detailed account of the spectral roc over the four vowel sections. In the voiceless context, the first section of the Wisconsin diphthong shows a higher spectral roc than the Ohio variant and reaches its spectral maximum in the second section while the Ohio /aɪ/ has its maximum later, in the third section. These differences most likely reflect dialect-specific articulatory movements and pace in order to attain the target offglide. In the voiced context, the spectral roc was much lower in general, and the maxima were reached later in the diphthong, in the fourth section of the Ohio vowel and in the third section of the Wisconsin variant. This temporal shift in the spectral roc maximum can be explained on the basis of the lengthening of the diphthongal onglide in the voiced context as discussed above. In general, then, the Wisconsin variant starts with faster articulatory movements which results in the higher roc; the Ohio variant begins with slower movements and reaches higher roc than the Wisconsin variant later in time, closer to its second target /ɪ/.

In sharp contrast to both Midwestern diphthongs is the spectral roc of the North Carolina /aɪ/ which did not change much over the course of its duration and increased only slightly in the fourth section. This result reflects the monophthongal property of /aɪ/ in this variety of American English.

C. Spectral dynamics and the listener

An obvious question arises as to how sensitive listeners are to these types of spectral changes. Do they recognize dialect-specific spectral variations or are these changes too minor to identify vowels as belonging or not belonging to their own regional variety? The spectral variations examined in this study were limited in terms of the type of consonantal context used and the number of levels of vowel emphasis. Clearly, other contexts will introduce other changes to the dynamic structure of vowels. Will the dialect-specific spectral features persist in these contexts or will they be obscured by contextual variation? These issues are important from yet another perspective, namely, that vowels change their properties across generations as a part of the process known as sound change. Will the dynamic spectral variations in vowels differ between younger and older speakers who grew up in the same dialect area? Will these variations be related to specific vowel shifts and changes currently taking place across geographic regions in the United States such as Northern Cities Shift or Southern Vowel shift? Further research is planned to determine whether and to what extent spectral dynamics contributes to the sound change in progress. The methods presented in this paper may prove useful in these efforts.

ACKNOWLEDGMENTS

This study was supported by Research Grant No. R01 DC006871 from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health. The authors would like to thank Joseph Salmons for his contributions to this research. The comments and suggestions of two anonymous reviewers on an earlier version of the paper are greatly appreciated.

[1]In this particular case, the acoustic proximity of both vowels may introduce some perceptual confusion, especially for listeners who grew up in a different dialect area. However, the differences in duration and in the position of the initial portion of each vowel including vowel nucleus (or 50% point) may contribute to a perceptual distinctiveness of both vowels.

[2]We used separate ANOVAs for each vowel rather than a single ANOVA because the main effect of vowel was expected to be significant for the present selection of vowels. Of interest to us were the dialectal differences within each vowel category and not the phonetic differences between vowels.

[3]For the vowels /ɪ, ɛ, æ/, the context × emphasis interaction arose from the fact that the spectral roc difference between the voiceless emphatic and voiceless nonemphatic contexts was significantly larger than that between voiced emphatic and voiced nonemphatic contexts. For /ɪ/, the mean values were 5.3 and 4.8 Hz/ms for voiceless emphatic vs nonemphatic context as compared to 4.6 and 4.7 Hz/ms for voiced emphatic vs nonemphatic, for /ɛ/: 5.1 and 4.2 Hz/ms vs 4.2 and 4.3 Hz/ms, respectively, and for /æ/: 4.1 and 3.8 Hz/ms vs 3.8 and 3.8 Hz/ms, respectively. However, the effects of context and emphasis created greater variation in the spectral roc for the diphthongal vowels /e, aɪ/. For /e/, it was the nonemphatic position in which roc was higher, and this was true for both voiceless and voiced contexts (the means were 6.1 and 6.6 Hz/ms for voiceless emphatic vs nonemphatic context as compared to 4.3 and 5.3 Hz/ms for voiced emphatic vs nonemphatic). For /aɪ/, there was a mixed pattern of variation in that the spectral roc was higher in the emphatic position for the voiceless context (means=9.8 and 9.1 Hz/ms for emphatic and nonemphatic, respectively) and it was higher in the nonemphatic position for the voiced (means=6.1 and 6.7 Hz/ms for emphatic and nonemphatic, respectively).

[4]Although we used statistical evidence as a guide to the strength of phonetic effects, we acknowledge complications in the interpretation of some of the present results. As shown in Lindblom *et al.* (2007), there is a complex interaction between variation as a function of consonantal context and emphatic stress so that "coarticulatory interactions between the C and V undergo complex, and often subtle, 'pulls' and 'pushes'" (p. 3803). In particular, characterization of locus equation slope (used as an index of degree of coarticulation) may be confounded by the effects of emphatic stress on F2 midpoint values. The effects of emphatic stress on the consonant onset and vowel F2 midpoint need to be separated. In the present study, the effects of emphatic stress may interact with the effects of consonantal context on vowel inherent spectral change in ways difficult to assess given the present design, in that we made only limited systematic modifications of phonetic context. Our current focus was to examine whether the effects of dialect on formant dynamics still persist in the presence of variation coming from the two phonetic sources.

Adank, P., van Hout, R., and Smits, R. (**2004**). "An acoustic description of the vowels of Northern and Southern Standard Dutch," J. Acoust. Soc. Am. **116**, 1729–1738.

Andruski, J. E., and Nearey, T. M. (**1992**). "On the sufficiency of compound target specification of isolated vowels and vowels in /bVb/ syllables," J. Acoust. Soc. Am. **91**, 390–410.

Byrd, D. (**1994**). "Relations of sex and dialect to reduction," Speech Commun. **15**, 39–54.

Clopper, C. G., Pisoni, D., and de Jong, K. (**2005**). "Acoustic characteristics of the vowel systems of six regional varieties of American English," J. Acoust. Soc. Am. **118**, 1661–1676.

Ferguson, S. H., and Kewley-Port, D. (**2002**). "Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **112**, 259–271.

Fox, R. A. (**1983**). "Perceptual structure of monophthongs and diphthongs in English," Lang Speech **26**, 21–60.

Gay, T. (**1968**). "Effect of speaking rate on diphthong formant movements," J. Acoust. Soc. Am. **44**, 1570–1573.

Harrington, J. (**2006**). "An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts," J. Phonetics **34**, 439–457.

Harrington, J., and Cassidy, S. (**1994**). "Dynamic and target theories of vowel classification: Evidence from monophthongs and diphthongs in Australian English," Lang Speech **37**, 357–373.

Harrington, J., Kleber, F., and Reubold, U. (**2008**). "Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study," J. Acoust. Soc. Am. **123**, 2825–2835.

Hillenbrand, M., and Gayvert, R. T. (**1993**). "Vowel classification based on fundamental frequency and formant frequencies," J. Speech Hear. Res. **36**, 694–700.

Hillenbrand, J. M., and Nearey, T. M. (**1999**). "Identification of resynthesized /hVd/ utterances: Effects of formant contour," J. Acoust. Soc. Am. **105**, 3509–3523.

Hillenbrand, J. M., Clark, M. J., and Nearey, T. M. (**2001**). "Effects of consonantal environment on vowel formant patterns," J. Acoust. Soc. Am. **109**, 748–763.

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

Jacewicz, E., Fox, R. A., and Salmons, J. (**2007**). "Vowel duration in three American English dialects," Am. Speech **82**, 367–385.

Jacewicz, E., Fox, R. A., O'Neill, K., and Salmons, J. (**2009**). "Articulation rate across dialect, age, and gender," Lang. Var. Change **21**, 233–256.

Jacewicz, E., Fujimura, O., and Fox, R. A. (**2003**). "Dynamics in diphthong perception," in Proceedings of the XVth International Congress of Phonetic Sciences, edited by M. J. Solé, D. Recasens and J. Romero, Barcelona, Spain, pp. 993–996.

Jenkins, J. J., Strange, W., and Edman, T. (**1983**). "Identification of vowels in 'vowelless' syllables," Percept. Psychophys. **34**, 441–450.

Jenkins, J. J., Strange, W., and Miranda, S. (**1994**). "Vowel identification in mixed-speaker silent-center syllables," J. Acoust. Soc. Am. **95**, 1030–

1043.
Kewley-Port, D., and Neel, A. (**1996**). "Perception of dynamic properties of speech: Peripheral and central processes," in *Listening to Speech: An Auditory Perspective*, edited by S. Greenberg and W. A. Aisnworth (Lawrence Erlbaum Associates, London), pp. 49–61.
Labov, W., Ash, S., and Boberg, C. (**2006**). *Atlas of North American English: Phonetics, Phonology, and Sound Change* (Mouton de Gruyter, Berlin).
Lehiste, I., and Peterson, G. (**1961**). "Transitions glides, and diphthongs," J. Acoust. Soc. Am. **33**, 268–277.
Lindblom, B. (**1963**). "Spectrographic study of vowel reduction," J. Acoust. Soc. Am. **35**, 1773–1781.
Lindblom, B., Agwuele, A., Sussman, H. M., and Cortes, E. E. (**2007**). "The effect of emphatic stress on consonant vowel coarticulation," J. Acoust. Soc. Am. **121**, 3802–3813.
Milenkovic, P. (**2003**). TF32 software program, University of Wisconsin, Madison, WI.
Moon, S.-J., and Lindblom, B. (**1994**). "Interaction between duration, context, and speaking style in English stressed vowels," J. Acoust. Soc. Am. **96**, 40–55.
Morrison, G. S. (**2009**). "Likelihood-ratio forensic voice comparison using parametric representations of the formant trajectories of diphthongs," J. Acoust. Soc. Am. **125**, 2387–2397.
Nearey, T. M., and Assmann, P. F. (**1986**). "Modeling the role of inherent spectral change in vowel identification," J. Acoust. Soc. Am. **80**, 1297–1308.
Olive, J. P., Greenwood, A., and Coleman, J. (**1993**). *Acoustics of American English Speech* (Springer-Verlag, New York).
Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**, 693–703.
Pickett, J. M. (**1999**). *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology* (Allyn and Bacon, Boston, MA).
Strange, W. (**1987**). "Information for vowels in formant transitions," J. Mem. Lang. **26**, 550–557.
Strange, W. (**1989**). "Dynamic specification of coarticulated vowels spoken in sentence context," J. Acoust. Soc. Am. **85**, 2135–2153.
Strange, W., Jenkins, J. J., and Johnson, T. L. (**1983**). "Dynamic specification of coarticulated vowels," J. Acoust. Soc. Am. **74**, 695–705.
Strange, W., Verbrugge, R. R., Shankweiler, D. P., and Edman, T. R. (**1976**). "Consonant environment specifies vowel identity," J. Acoust. Soc. Am. **60**, 213–224.
Van Son, R. J. J. H., and Pols, L. C. W. (**1992**). "Formant movements of Dutch vowels in a text, read at normal and fast rate," J. Acoust. Soc. Am. **92**, 121–127.
Watson, C. I., and Harrington, J. (**1999**). "Acoustic evidence for dynamic formant trajectories in Australian English vowels," J. Acoust. Soc. Am. **106**, 458–468.
Wouters, J., and Macon, M. W. (**2002a**). "Effects of prosodic factors on spectral dynamics, I. Analysis," J. Acoust. Soc. Am. **111**, 417–427.
Wouters, J., and Macon, M. W. (**2002b**). "Effects of prosodic factors on spectral dynamics, II. Synthesis," J. Acoust. Soc. Am. **111**, 428–438.
Zahorian, S. A., and Jagharghi, A. (**1993**). "Spectral-shape features versus formants as acoustic correlates for vowels," J. Acoust. Soc. Am. **94**, 1966–1982.

Acoustic measurement of overall voice quality: A meta-analysis[a)]

Youri Maryn[b)]
Department of Speech-Language Pathology and Audiology, Sint-Jan General Hospital, Ruddershove 10, 8000 Bruges, Belgium; Faculty of Health Care Vesalius, University College Ghent, Keramiekstraat 80, 9000 Ghent, Belgium; and Faculty of Social Health Sciences, University of Ghent, De Pintelaan 185, 9000 Ghent, Belgium

Nelson Roy
Department of Communication Sciences and Disorders, Division of Otolaryngology—Head and Neck Surgery, The University of Utah, 390 South 1530 East, Salt Lake City, Utah 84112-0252

Marc De Bodt
Department of Communication Disorders, University Hospital of Antwerp, Wilrijkstraat 10, 2650 Edegem, Belgium and Faculty of Social Health Sciences, University of Ghent, De Pintelaan 185, 9000 Ghent, Belgium

Paul Van Cauwenberge
Faculty of Medicine and Social Health Sciences, University of Ghent, De Pintelaan 185, 9000 Ghent, Belgium

Paul Corthals
Faculty of Health Care Vesalius, University College Ghent, Keramiekstraat 80, 9000 Ghent, Belgium and Faculty of Social Health Sciences, University of Ghent, De Pintelaan 185, 9000 Ghent, Belgium

(Received 27 July 2008; revised 10 August 2009; accepted 12 August 2009)

Over the past several decades, many acoustic markers have been proposed to be sensitive to and measure overall voice quality. This meta-analysis presents a retrospective appraisal of scientific reports, which evaluated the relation between perceived overall voice quality and several acoustic-phonetic correlates. Twenty-five studies met the inclusion criteria and were evaluated using meta-analytic techniques. Correlation coefficients between perceptual judgments and acoustic measures were computed. Where more than one correlation coefficient for a specific acoustic marker was available, a weighted average correlation coefficient was calculated. This was the case in 36 acoustic measures on sustained vowels and in 3 measures on continuous speech. Acoustic measures were ranked according to the strength of the correlation with perceptual voice quality ratings. Acoustic markers with more than one correlation value available in literature and yielding a homogeneous weighted r of 0.60 or above were considered to be superior. The meta-analysis identified four measures that met these criteria in sustained vowels and three measures in continuous speech. Although acoustic measures are routinely utilized in clinical voice examinations, the results of this meta-analysis suggest that caution is warranted regarding the concurrent validity and thus the clinical utility of many of these measures. © *2009 Acoustical Society of America.*
[DOI: 10.1121/1.3224706]

PACS number(s): 43.70.Jt, 43.72.Ar [AL] Pages: 2619–2634

I. INTRODUCTION

Evaluation of voice quality is considered an essential but controversial part of the assessment process in the field of voice pathology. In clinical as well as in research settings, two main approaches exist to describe the perceived severity of a voice disorder (Kreiman and Gerratt, 2000a). First, generic and/or global ratings such as "overall voice quality," also known as "G" (for "grade"), "severity of voice disorder," "severity of dysphonia," "overall abnormality," and "overall severity" have been used to capture a composite perceptual judgment of the degree of the perceived dysphonia. In contrast, other voice quality ratings pertain to single and very specific perceptual dimensions, the best known of which are roughness and breathiness. Recent evidence has suggested that perceptual rating of overall voice quality and other more specific perceptual dimensions is difficult, as such judgments depend on the listener's internal standard or scale for voice quality dimensions, on his/her sensitivity for this particular dimension, on fatigue, attention, exposure to various disordered voices, and training in perceptual evaluation of voice quality (Kreiman *et al.*, 1993; Eadie and Baylor, 2006). Furthermore, other aspects of voice quality judgments, such as type and range of the scale (Bele, 2005; Eadie and Doyle, 2002), or the type of sample to be evaluated, such

[a)]Portions of this work were presented in "Forty years of acoustic prediction of overall voice quality" at the 7th International Voice Symposium, Austrian Voice Institute, Salzburg, Austria, August 4–6, 2006.
[b)]Author to whom correspondence should be addressed. Electronic mail: youri.maryn@azbrugge.be

as sustained vowel versus continuous speech (Bele, 2005; Zraick *et al.*, 2005; Eadie and Baylor, 2006), can significantly affect the perceptual evaluation of voice quality.

In spite of these listener-related and other potential biases, many researchers have tried to correlate the outcome of acoustic-phonetic measures to vocal quality ratings and dysphonia severity. The replacement of analog recording systems with digital recording systems, the availability of automated analysis algorithms, and the non-invasiveness of acoustic measures, combined with the fact that acoustic parameters provide easy quantification of dysphonia improvement during the treatment process, have led to considerable interest in clinical voice quality measurement using acoustic analysis techniques.

In this regard, the correlation coefficient has emerged as the most frequently used index to determine the extent of the relationship or effect size between acoustic measures and listener judgments of dysphonia severity. The correlation coefficient as a measure of effect size measures the strength and direction of a linear relationship between two variables. In the voice quality literature, perceived overall voice quality is treated as the dependent variable with the objective acoustic measure treated as the independent variable. The degree of the linear relationship between dependent and independent variables (i.e., correlation) counts as an indication of validity, or the extent to which the score of a measurement (i.e., the acoustic parameter) can be regarded as a valid measure of the dependent variable (i.e., the perceptual rating). Consequently, the higher the absolute correlation coefficient, the more the acoustic measure is said to reflect the perception of overall voice quality, and vice versa. The correlation coefficient is thus an important and frequently used statistic in voice quality research, especially to validate acoustic measures.

Although the correlation coefficient is a frequent metric to assess the strength of the acoustic-perceptual relationship, at least 60 possible acoustic determinants of overall voice quality with varying predictive power have been identified in literature over the past 4 decades. Buder (2000) proposed a taxonomy of 15 signal processing-based categories to help manage the wide array of acoustic measures. The large numbers of studies reviewed by Buder (2000) clearly differ substantially in the number of participants and the magnitude of correlation with perceptual judgments of voice quality. Furthermore, the signal processing strategies vary from classic spectrography to sophisticated statistics on sound wave microstructure. Whereas some authors examined the predictive power of resonance-based aspects, the majority of investigators focused on glottal rather than on supraglottal phenomena, seeking correlates of overall voice quality in the distribution of fundamental frequency, in waveform perturbations, in various spectral parameters (including cepstral coefficients and noise content of the glottal sound source), in glottal air flow models obtained by inverse filtering, or in models based on non-linear dynamics theory.

Although an impressive body of research exists, which ostensibly assesses the utility of acoustic measurement to quantify voice quality and dysphonia severity, procedural differences in type and number of acoustic predictors, type of recorded material, analysis equipment, and measurement scales have made it almost impossible to qualitatively appraise the merits of these studies, and precisely define a subset of the most robust and sensitive acoustic measures. One approach to this seemingly intractable problem is to apply meta-analytic techniques. Meta-analysis refers to "the analysis of analyses," and is a statistical technique for amalgamating, summarizing, and reviewing previous quantitative research. Unlike traditional research methods, meta-analysis uses the summary statistics from individual studies as the data points for the purpose of integrating the findings. A key assumption of this analysis approach is that each study provides a different estimate of the underlying relationship within the population. By accumulating results across studies, one can gain a more accurate representation of the population relationship than is provided by the individual study estimators. In this way, meta-analyses permit confidence that the reported results are based on more than one study that found the same result (Frey *et al.*, 1991; Lipsey and Wilson, 2001).

Meta-analysis reviews findings in terms of effect sizes. Defining an effect size statistic that adequately represents the quantitative findings of an assortment of research reports in a standardized profile is essential to meta-analysis, as it permits meaningful numerical comparison and analysis (Lipsey and Wilson, 2001). The effect size provides information about the magnitude of the relationships observed across all studies and for subsets of studies. By treating individual correlation coefficients as indicators of effect size, meta-analysis can regroup study outcomes into homogeneous subsets and establish population effect sizes. The population effect size, i.e., the real relationship between an independent variable (a specific acoustic measure) and the dependent variable (a voice severity rating), is estimated by a "weighted" average of all correlations available for a particular acoustic marker. In addition to defining a weighted average of all effect sizes (i.e., correlation coefficients) in a meta-analysis, it is also important to know whether or not the various effect sizes all estimate the same population effect size. This is a question of homogeneity (or heterogeneity) of the effect size distribution, and a population effect size can only be interpreted reliably if the underlying data set is sufficiently homogeneous (Hunter *et al.*, 1982). When the variability of effect sizes around their weighted mean is no larger than the dispersion expected from sampling error alone, the effect size distribution is considered to be homogeneous. By comparison, in a heterogeneous distribution, individual effect sizes differ from the weighted mean by more than the sampling error (Lipsey and Wilson, 2001). Multiple correlation coefficients resulting in a homogeneous weighted mean correlation are considered to confirm each other, thereby increasing the generalizability of the findings.

Given the large body of research that relates acoustic measures to voice quality ratings, meta-analysis techniques can potentially reduce information overload, and distill this large literature into a manageable and/or tractable set of conclusions. Therefore, the aim of this meta-analysis is twofold: (1) to retrospectively appraise the acoustic-phonetic markers

for overall voice quality (e.g., dysphonia severity) and (2) to establish population relationship estimates for several acoustic measures.

II. METHOD

In most research on assessment of voice quality, measurements have been completed on sustained vowels as compared to continuous speech. This preference for sustained vowels over continuous speech in acoustic as well as perceptual measurements of voice quality has been motivated by several factors (Askenfelt and Hammarberg, 1986; Parsa and Jamieson, 2001), such as follows: (a) sustained vowels represent relatively time-invariant vocal phonation whereas continuous speech involves quick and continuous alterations of glottal and supraglottal mechanisms; (b) in contrast to continuous speech, sustained mid-vowel segments do not contain non-voiced phonemes, rapid voice onsets and offsets, or prosodic fundamental frequency and amplitude fluctuations; and (c) sustained vowels are not affected by speech rate, vocal pauses, phonetic context, and stress. However, sustained vowels may lack representation of daily speech and voice (Parsa and Jamieson, 2001; Eadie and Baylor, 2006), and continuous speech potentially contains perceptual cues, which are often considered to be decisive in vocal quality evaluations (Askenfelt and Hammarberg, 1986). Since both sample types offer valuable information in voice quality measurements, the present meta-analysis focused on studies of sustained vowels as well as connected speech.

A. Search strategy

Relevant scientific reports were identified by a systematic electronic search of the Medline database and the corpus of online publications by the American Speech-Language-Hearing Association. The combination of (a) keywords referring to composite perceptual voice evaluations and (b) keywords related to the concepts of measuring by means of acoustic markers was used as a guide. Using information derived from the titles and abstracts, an initial set of pertinent articles was generated. Subsequently, a manual search for references in relevant literature sources was launched using the same guide. This manual search started from the sources cited in the initial set of articles garnered from the electronic search and from periodicals, book chapters, and various bibliographies likely to contain relevant references and texts.

B. Inclusion and exclusion of literature sources

In order to be included, a study had to report sufficient mathematical detail on bivariate correlation coefficients establishing the relation between perceptual overall voice quality ratings of sustained vowels or continuous speech (the dependent variable) and one or more acoustic parameters derived from the same samples (the independent variables). Studies citing relevant correlation coefficients were included, whether or not significance levels were reported, and every study describing auditory-perceptual ratings of overall quality (e.g., dysphonia severity) was included, regardless of the type of rating scale used.

Investigations of acoustic correlates of specific perceptual dimensions such as breathiness and roughness were not included in the meta-analysis, as the present study concentrated on "composite" or "global" overall voice quality correlates. Furthermore, reports on non-acoustic or non-objective correlates, such as aerodynamic measures or electroglottographic parameters, were also excluded, as well as studies dealing with the relationship between the auditory-perceptual evaluation of overall voice quality and its visual-perceptual representation in narrowband spectrograms. Furthermore, since the present study aimed to focus on acoustic-auditory determinants of dysphonia severity, studies investigating the correlation between objective acoustic measures and visual inspection of spectrograms or other diagrams were excluded. Also, reports on parameters derived from synthesized vowel samples were not included in this study. Reports lacking sufficient quantitative and critical information, such as number of subjects or type of samples, were also excluded.

Methodological articles related exclusively to the use and development of perceptual rating scales or acoustic algorithms, which did not provide inferential statistics on the validity of the acoustic measure(s), were also excluded. Studies appraising the diagnostic value of acoustic parameters (i.e., the power of a diagnostic tool to discriminate between presence or absence of a voice disorder), expressed as sensitivity, specificity, positive predictive value, negative predictive, and/or area under the receiver operating characteristic (ROC) curve, or outcomes of studies based on comparative statistics between normal and pathologic voices, as expressed in chi-square tests, Mann-Whitney U tests, t tests, etc., were not included, because the present study concentrated on the correlation coefficient as population effect size.

In addition, reports on multivariate analyses were excluded, unless bivariate (zero-order) correlation coefficients were clearly identified (as in Wolfe and Martin, 1997; Wolfe *et al.*, 1997; Yu *et al.*, 2001; Eadie and Baylor, 2006; Ma and Yiu, 2006). The reason for excluding multiple regression studies is based on the assumption that some independent variables are dropped from the initial set of possible predictors as a result of co-linearity. A relevant independent variable, correlating well with the dependent variable, may be dropped when, in the presence of other independent variables, it does not substantially increase the amount of variance explained. This phenomenon makes it difficult to assess the separate contribution of each independent variable to the measurement of the dependent variable. Moreover, the algorithm of a multiple regression not only looks for a parsimonious equation, but also gives each remaining independent marker a coefficient that can only be interpreted in combination with the particular set of remaining independent markers in the rest of the equation. As a consequence, we had to exclude study results using multivariate statistics from the meta-analysis.

Finally, reliability of the auditory-perceptual ratings of voice quality, as an index on which an acoustic measure is validated, is also an important consideration (Kreiman and Gerratt, 2000a, 2000b; Kreiman *et al.*, 2007). Reliability of auditory-perceptual ratings is traditionally described in terms

of within and between listener reliability, consistency, agreement, or concordance. Such intra- and interrater reliability is considered an important prerequisite for validity. High reliability clearly and precisely defines the perceptual construct to be measured by an acoustic parameter. In contrast, listener unreliability increases "non-experimental" or "error" variance, thereby reducing the true variance in the perceptual construct that is to be accounted for by the acoustic measure. Thus, the increase in error variance due to listener unreliability should decrease the concurrent validity of the acoustic measure, as evaluated by a correlation coefficient. In single experiments, acceptable rater reliability is often considered an essential prerequisite before attempting to assess an acoustic measure's worth in estimating overall voice quality. However, across studies, many statistics have been used to measure rater reliability, for instance, Pearson's product-moment correlation coefficient, Cohen's kappa correlation coefficient, Cronbach's alpha correlation coefficient, and intraclass correlation coefficient, to mention only a few. Given the large number of studies reviewed in this meta-analysis, each using a variety of listeners (with differing levels of experience and training), and different scales with various interpretation guidelines so as to determine "adequate" reliability, we elected to treat listener reliability as a nuisance variable, and to not exclude any studies solely on the basis of their estimates of listener reliability. This decision is predicated on the assumption that listener unreliability essentially contributes to error variance, and necessarily attenuates any investigator's ability to identify significant correlations between listener ratings and specific acoustic measures. By treating listener reliability/unreliability as a nuisance variable, one that would necessarily vary between studies and differentially contribute to error variance, we assumed that across studies, the most compelling acoustic-perceptual relationships would eventually surface, having survived the potentially attenuating effects of listener unreliability.

Analogous to the listener reliability/unreliability, we also elected to treat between-study differences in data acquisition and processing methodology as a nuisance variable. Variety in room acoustics, microphone type and placement, software, analysis algorithms, etc., also creates error variance, and similarly decreases the variance in the perceptual construct that is to be explained by the acoustic measure. The large number of studies, each with its own acoustical configuration and hardware and software settings, clearly limits our ability to directly compare the outcomes of the studies. However, a guiding principle of meta-analysis is that the consistency of the significant results/conclusions across studies is paramount, and robust relationships should withstand such methodological "noise" (regardless of the source of the noise, e.g., listener unreliability, recording instrumentation and surroundings, computer software, etc). We therefore elected to consider methodological variations in recording conditions/settings, data acquisition and analysis algorithms, etc., as additional sources of error variance, and an inherent limitation of the meta-analysis. On the other hand, acoustic measures that yield consistent outcomes across a variety of study methods can be considered especially robust. As such, the inclusion of studies with varying methodology is considered advantageous in the present meta-analysis.

Originally, 85 reports were considered. Based on the aforementioned inclusion and exclusion criteria, however, many reports were excluded, producing a final corpus of 25 studies on which the meta-analysis was performed. Twenty-one studies involved measurements on sustained vowel samples (methodological aspects of these studies are summarized in Table I). Seven studies involved measurement on continuous speech samples (methodological aspects of these studies are similarly summarized in Table II). However, 3 studies contained information on both continuous speech and sustained vowels (Heman-Ackah *et al.*, 2002; Halberstam, 2004; Eadie and Baylor, 2006), thus leaving a total of 25 studies (i.e., $28-3=25$).

From these studies, a list of acoustic measures was generated. Subsequently, the measures were organized based on their description in the Method section of the original publication. The tabulation of Buder (2000) was chosen only as a loose framework to group the measures. Buder's (2000) tabulation was the first compilation of acoustic voice measures, as it presented a complete overview of acoustic measures in a comprehensive and consistently structured manner. It therefore served as a basis on which the measures of this meta-analysis were considered to be similar or different. In studies that analyzed sustained vowels, there were 69 acoustic markers identified, whereas in the connected speech studies, 26 acoustic measures were reported. Eighty-seven acoustic markers have been identified as measures of overall voice quality in the included studies. Table III lists these acoustic measures alphabetically and provides (for every measure) references expanding on the rationale and the digital signal processing underlying the measure.

C. Statistical analysis

Quantitative data from the selected scientific reports were analyzed using statistical software packages for personal computers, including Microsoft Office Excel 2003 and Meta-Analysis Programs version 5.3 (Ralf Schwarzer, Department of Psychology, Freie Universität Berlin, Germany). Meta-analyses on correlation coefficients according to the Schmidt-Hunter method (Hunter *et al.*, 1982) were performed on all acoustic voice quality correlates for which more than one effect size was available. This method is based on four statistics. The first statistic is the *number of effect sizes* (k) or the number of available bivariate correlation coefficients for a given acoustic measure. The second statistic is the total *number of subjects* (N). The third statistic is the *population effect size* or the weighted mean correlation coefficient ($\overline{r_w}$). Correlation coefficients based on studies with large sample sizes digress less from the population effect size and therefore more weight is assigned to large N effect sizes (Hunter *et al.*, 1982). If only one effect size is reported, a weighted effect size cannot be calculated. In this case, there is no meta-analysis and the discussion is based on the initial and solitary r-value. While there is no firm criterion or universal consensus for evaluating the magnitude of correlation coefficients (Frey *et al.*, 1991), we chose a corre-

TABLE I. Methodological features (number of subjects, type of voice recording, and organization and reliability of the perceptual ratings) of the 21 chronologically ordered studies included in this meta-analysis on sustained vowels.

Source	Subjects[a]			Voice sample[b]		Perceptual evaluation[b]				
	N	*P*	*T*	Vowel	Duration	No. of judges	Rating scale[c]	Perceptual construct	Intrarater reliability[d]	Interrater reliability[d]
Kojima *et al.* (1980)	28	30	58	/a/	NA	5	EAI (4)	Hoarseness	NA	NA
Yumoto *et al.* (1984)	0	87	87	/a/	3 s	8	EAI (4)	Hoarseness	NA	0.51–0.79 Sp
Hirano *et al.* (1986)	0	68	68	/e/	NA	NA	EAI (4)	G, grade	NA	NA
Prosek *et al.* (1987)	0	90	90	/a/	2 s	9	EAI (7)	Severity of voice disorder	0.90 Pe	0.82 Cr
	16	44	60			14		Hoarseness	NA	NA
Wolfe and Steinfatt (1987)	0	51	51	/a/ and /i/	1 s	8	EAI (7)	Severity of dysphonia Hoarseness	89% Ag	0.95 Cr
Feijoo and Hernández (1990)	64	57	121	/e/	NA	4	EAI (4)	G, grade	77.48% Ag	98.35% Ag
Kreiman *et al.* (1990)	0	18	18	/a/	1.67 s	10	EAI (7)	Overall abnormality	NA	NA
Wolfe *et al.* (1995)	20	60	80	/a/	1 s	22	EAI (7)	Overall severity	0.99 Cr	0.98 Cr
Dejonckere *et al.* (1996)	0	943	943	/a/	2 s	2	EAI (4)	G, grade	0.51 Co	0.87 Sp
Dejonckere and Wieneke (1996)	0	28	28	/a/	0.1 s	2	EAI (5)	Overall severity of hoarseness	NA	NA
De Bodt (1997)	98	634	732	/a/	3 s	1	EAI (4)	G, grade	NA	NA
Plant *et al.* (1997)	0	26	26	/i/	2 s	3	EAI (5)	Overall voice quality	0.86 NA	NA
Wolfe and Martin (1997)	0	51	51	/a/ and /i/	1 s	11	EAI (7)	Hoarseness	76% Ag	0.95 Cr
Wolfe *et al.* (1997)	0	51	51	/a/ and /i/	1 s	18	VAS	Hoarseness	0.80 Pe	0.94 Cr
Wolfe *et al.* (2000)	0	20	20	/a/	1 s	11	EAI (7)	Abnormality	0.81 Pe	0.98 Cr
Yu *et al.* (2001)	21	63	84	/a/	2 s	6	EAI (4)	G, grade	Cons	Cons
Heman-Ackah *et al.* (2002)	0	14	14	/a/	1 s	2	EAI (4)	G, grade	NA	0.83 Pe
Halberstam (2004)	0	60	60	/a/	1 s	2	EAI (7)	Hoarseness	0.89 NA	0.91 Cr
Eadie and Baylor (2006)	3	9	12	/a/	1 s	16	VAS	Overall severity	0.82–0.95 Pe	0.72–0.83 Pe
Gorham-Rowan and Laures-Gore (2006)	0	28^{ym}	28^{ym}	/a/	1 s	10	FMMEP	Hoarseness	−0.32 to 0.86 Pe	0.80 Cr
	0	28^{ew}	28^{ew}							
	0	28^{em}	28^{em}							
Yu *et al.* (2007)	38^{w}	270^{w}	308^{w}	/a/	2 s	4	VAS	G, grade	NA	NA
	20^{m}	121^{m}	141^{m}							

[a]*N*=number of normal subjects, *P*=number of pathological or dysphonic subjects, *T*=total number of subjects, ^{ym}=young men, ^{ew}=elderly women, ^{em}=elderly men, ^{m}=men, and ^{w}=women.

[b]NA=the information is not available in the original manuscript.

[c]EAI=equal-appearing interval scale with between parentheses the number of points on the scale, VAS=visual analog scale, and FMMEP=free modulus magnitude estimation paradigm.

[d]Sp=Spearman's rank-order correlation coefficient, Pe=Pearson's product-moment correlation coefficient, Co=Cohen's kappa correlation coefficient, Cr=Cronbach's alpha correlation coefficient, Ke=Kendall's coefficient of concordance, Ag=percentage of agreement/consistency between judgments, and Cons=consensus between listeners without quantitative measure of reliability.

lation coefficient (r or $\overline{r_w}$) of 0.60 as the cutoff to distinguish between strong and weak acoustic markers. Following the guidelines established by Franzblau (1958), this threshold intends to separate a "moderate" degree of correlation from a "marked" degree of correlation. It should be acknowledged, however, that other interpretations have been proposed, including Frey *et al.* (1991), for example, who recommended r of 0.70 to distinguish between moderate and marked correlations. We selected a less stringent correlation coefficient $r=0.60$ in light of our decision to treat listener unreliability and methodological/procedural differences as sources of error variance, which would potentially attenuate the strength of reported bivariate correlations across studies. The fourth statistic relates to the homogeneity or heterogeneity of the effect sizes. A population effect size can only be interpreted reliably if the underlying data set is sufficiently homogeneous (Hunter *et al.*, 1982). Here one can rely on several indicators: (1) the residual standard deviation, (2) the percentage of observed variance accounted for by the sampling error, and (3) the chi-square value. However, the preferred index for homogeneity is the population variance or its square root, called *residual standard deviation* (SD_{res}). This indicator, SD_{res}, is the variance left after the sampling error has been subtracted (Hunter *et al.*, 1982). Ideally, SD_{res} equals zero, meaning that all the observed variance is accounted for by sampling error and that the data set of correlations is completely homogeneous. If the analysis, however, failed to identify a source of systematic variation in the data, SD_{res} is indicative of heterogeneity. As a rule of thumb, a set of effect sizes can be considered homogeneous when SD_{res} is less than $\frac{1}{4}$ of $\overline{r_w}$ (Hunter *et al.*, 1982; Lipsey and Wilson, 2001).

III. RESULTS

A. Sustained vowels

Twenty-one studies meeting the selection criteria were identified; the majority originated from the *Journal of*

TABLE II. Methodological features (number of subjects, type of voice recording, and organization and reliability of the perceptual ratings) of the seven chronologically ordered studies included in this meta-analysis on continuous speech.

Source	Subjects[a]			Voice sample	Perceptual evaluation[b]				
	N	*P*	*T*		No. of judges	Rating scale[c]	Perceptual construct	Intrarater reliability[d]	Interrater reliability[d]
Askenfelt and Hammarberg (1986)	0	41	41	Voiced segments, 40 s of reading a story	6	EAI (6)	Overall voice quality	0.86–0.98 Pe	NA
Qi *et al.* (1999)	0	87	87	First and second sentences from rainbow passage	5	VAS	Overall voice quality	0.93–0.96 Pe	0.97 Cr
Heman-Ackah *et al.* (2002)	0	18	18	Second sentence from rainbow passage	2	EAI (4)	G, grade	NA	0.83 Pe
Halberstam (2004)	0	60	60	12 s from rainbow passage	2	EAI (7)	Hoarseness	0.93 NA	0.97 Cr
Eadie and Doyle (2005)	6	24	30	Second sentence from rainbow passage	12	DME	Overall severity	0.69 Pe	0.97 Cr
Eadie and Baylor (2006)	3	9	12	Second sentence from rainbow passage	16	VAS	Overall severity	0.80–0.97 Pe	0.84–0.91 Pe
Ma and Yiu (2006)	41	112	153	/ba ba d a bo /	4	EAI (11)	G, grade	≥0.90 Pe	0.86–0.91 Pe

[a]*N*=number of normal subjects, *P*=number of pathological or dysphonic subjects, and *T*=total number of subjects.
[b]NA=the information is not available in the original manuscript.
[c]EAI=equal-appearing interval scale with between parentheses the number of points on the scale, VAS=visual analog scale, and DME=direct magnitude estimation.
[d]Pe=Pearson's product-moment correlation coefficient and Cr=Cronbach's alpha correlation coefficient.

Speech (Language) and Hearing Disorders (6), *Journal of Voice* (3), and *Journal of Communication Disorders* (3). Other sources were *Acta Otorhinolaryngologica Belgica* (1), *Acta Otolaryngologica* (1), *Folia Phoniatrica et Logopaedica* (2), *Journal of Phonetics* (1), *Laryngoscope* (1), *ORL* (1), *Revue de Laryngologie-Otologie-Rhinologie* (1), and a chapter in volume VI of *Advances in Clinical Phonetics* (1). Relevant information concerning the methodology of the reports that were included in the meta-analysis can be found in Table I. All 21 studies reported on pathologic or dysphonic voices; however, only 8 studies also contained normal voices. The mean number of dysphonic voice samples was 115 (range 9–943). For the normal voices, the mean number was 34 (range 3–98). The total number of subjects was 116 on average and ranged from 12 to 943. In these studies, 146 distinct effect sizes (i.e., correlation coefficients) were reported, pertaining to 69 different acoustic measures as displayed in Table IV.

All acoustic parameters and data on sustained vowels were extracted from the central portion of the recordings. The length of the mid-vowel segment varied from 0.1 to 3 s with a mean duration of 1.5 s. 1 s was the modal duration, occurring in 50% of the studies (the duration was not specified in three studies). The vowels [a:], [i:], and [e:] were analyzed in 86%, 19%, and 10% of the studies, respectively. Substantial differences existed among the data acquisition systems that were used, which could potentially influence the outcome of acoustic measurements. For instance, recording equipment (e.g., type of microphone and microphone location relative to the sound source, and type of hardware), processing algorithms, measurement algorithms, and software settings such as sampling rate or method of fundamental period extraction varied among the studies and have been demonstrated to influence the outcome of acoustic measurements, particularly the outcomes of perturbation measures.

For the perceptual experiments, the number of judges ranged from 1 to 22, with a mean value of 8. The rating scale used was typically an equal-appearing interval scale, using 4, 5, or 7 points in 38%, 10%, and 33% of studies, respectively. In two studies (Wolfe *et al.*, 1997; Yu *et al.*, 2007), a visual analog scale was used. In Gorham-Rowan and Laures-Gore, 2006 a free modulus magnitude estimation paradigm was used. A variety of perceptual labels were used including hoarseness, G (from grade), severity of voice disorder, severity of dysphonia, overall abnormality, overall severity, overall severity of hoarseness, abnormality, and overall voice quality. A variety of estimates of inter- and intrajudge reliability estimates were used (see table entries). As mentioned previously, the variety and range of methods to determine reliability hamper comparisons between studies. In general, intrajudge reliability fluctuated from rather low, as in Dejonckere and Wieneke, 1996 and Gorham-Rowan and Laures-Gore, 2006, to very high, as in Wolfe *et al.*, 1995 and Prosek *et al.*, 1987. Similar variability was observed for interjudge reliability.

1. Meta-analysis on correlation coefficients

The results of the meta-analysis on sustained vowels are summarized in Table IV and Fig. 1. For 33 of the 69 acoustics measures (48%), there was only 1 correlation coefficient available and, consequently, no weighted mean correlation coefficient could be determined. For the remaining 36 acoustic determinants (52%), there was more than 1 correlation coefficient and the *k*-values ranged from 2 to 7. The most frequently investigated parameters were noise-to-harmonics ratio (NHR) from multi-dimensional voice program (MDVP) ($k=7$), and the vocal perturbation measures amplitude perturbation quotient, percent jitter, and percent shimmer ($k=6$). For these 36 markers, a $\overline{r_w}$ was calculated with Meta-Analysis Programs version 5.3. The organization of the meta-analysis on acoustic measures on sustained vowels is illustrated in Fig. 1.

TABLE III. The 87 acoustic measures included in this study, alphabetically ordered on the basis of their full name, with their respective sources/citations identified.

Acoustic measure	Sources included in study
Absolute jitter	Kreiman *et al.* (1990), De Bodt (1997), Wolfe *et al.* (1997), and Halberstam (2004)
Amplitude perturbation quotient	De Bodt (1997), Wolfe *et al.* (1997), Heman-Ackah *et al.* (2002), Halberstam (2004), and Gorham-Rowan and Laures-Gore (2006)
Amplitude perturbation quotient of residue signal	Prosek *et al.* (1987)
Area of voice range profile	Ma and Yiu (2006)
Breathiness index	Plant *et al.* (1997) and Wolfe *et al.* (2000)
Cepstral peak magnitude (also known as magnitude of first harmonic)	Dejonckere and Wieneke (1996)
Cepstral peak prominence	Wolfe and Martin (1997), Wolfe *et al.* (2000), Halberstam (2004), and Eadie and Baylor (2006)
Cepstrum of excitation signal	Feijoo and Hernández (1990)
Coefficient of excess	Prosek *et al.* (1987)
Coefficient of variation of fundamental frequency	Wolfe *et al.* (1997)
Coefficient of variation of jitter	Kreiman *et al.* (1990)
Coefficient of variation of period	Wolfe and Steinfatt (1987)
Coefficient of variation of shimmer	Kreiman *et al.* (1990)
Compression of relative frequency differences	Askenfelt and Hammarberg (1986)
Cycle-of-cycle variation of waveform	Feijoo and Hernández (1990)
Difference between frequencies of second and first formants	Kreiman *et al.* (1990)
Directional perturbation factor	Askenfelt and Hammarberg (1986)
Fluctuation in amplitude	Hirano *et al.* (1986)
Fluctuation in fundamental frequency	Hirano *et al.* (1986)
Frequency-domain harmonics-to-noise ratio	Eadie and Doyle (2005)
Frequency of first formant	Kreiman *et al.* (1990)
Frequency of second formant	Kreiman *et al.* (1990)
Frequency of third formant	Kreiman *et al.* (1990)
Fundamental frequency	Yu *et al.* (2001), Yu *et al.* (2007), and Ma and Yiu (2006)
Fundamental frequency range in voice range profile	Ma and Yiu (2006)
Harmonics-to-noise ratio from Kojima	Kojima *et al.* (1980)
Harmonics-to-noise ratio from Yumoto	Yumoto *et al.* (1984), Kreiman *et al.* (1990), and Wolfe *et al.* (1995)
Highest fundamental frequency in voice range profile	Ma and Yiu (2006)
Intensity range in voice range profile	Ma and Yiu (2006)
Jitter factor	Yu *et al.* (2001) and Yu *et al.* (2007)
Jitter from Yumoto	Yumoto *et al.* (1984)
Jitter ratio	Wolfe and Steinfatt (1987) and Dejonckere and Wieneke (1996)
Lowest fundamental frequency in voice range profile	Ma and Yiu (2006)
Lyapunov coefficient	Yu *et al.* (2001) and Yu *et al.* (2007)
Maximum intensity in voice range profile	Ma and Yiu (2006)
Mean harmonic emergence between 500 and 1500 Hz	Dejonckere and Wieneke (1996)
Minimum intensity in voice range profile	Ma and Yiu (2006)
Natural logarithm of standard deviation of period	Wolfe and Steinfatt (1987) and Kreiman *et al.* (1990)
Noise-to-harmonics ratio	Wolfe *et al.* (1997)
Noise-to-harmonics ratio from MDVP	Dejonckere *et al.* (1996), De Bodt (1997), Heman-Ackah *et al.* (2002), Halberstam (2004), Gorham-Rowan and Laures-Gore (2006), and Ma and Yiu (2006)
Normalized mean absolute period jitter	Feijoo and Hernández (1990)
Normalized mean absolute period shimmer	Feijoo and Hernández (1990)
Normalized noise energy	Feijoo and Hernández (1990)
No. of harmonics	Kreiman *et al.* (1990)
Partial period comparison	Kreiman *et al.* (1990)
Peakedness of relative frequency differences	Askenfelt and Hammarberg (1986)
Pearson r at autocorrelation peak	Wolfe *et al.* (2000)
Percent jitter	Kreiman *et al.* (1990), De Bodt (1997), Plant *et al.* (1997), Wolfe and Martin (1997), Wolfe *et al.* (1997), and Halberstam (2004)
Percent shimmer	Kreiman *et al.* (1990), Dejonckere *et al.* (1996), De Bodt (1997), Wolfe and Martin (1997), Wolfe *et al.* (1997), Halberstam (2004), and Ma and Yiu (2006)
Perturbation factor	Askenfelt and Hammarberg (1986)

TABLE III. (*Continued.*)

Acoustic measure	Sources included in study
Perturbation magnitude	Askenfelt and Hammarberg (1986)
Perturbation magnitude mean	Askenfelt and Hammarberg (1986)
Phonatory fundamental frequency range	De Bodt (1997), Yu *et al.* (2001), Halberstam (2004), and Yu *et al.* (2007)
Pitch amplitude	Prosek *et al.* (1987), Plant *et al.* (1997), and Eadie and Doyle (2005)
Pitch perturbation quotient	De Bodt (1997), Wolfe *et al.* (1997), and Halberstam (2004)
Pitch perturbation quotient of residue signal	Prosek *et al.* (1987)
Power spectrum ratio	Wolfe *et al.* (2000)
Ratio of amplitudes of first and second harmonics	Kreiman *et al.* (1990)
Ratio of frequencies of second and first formants	Kreiman *et al.* (1990)
Relative average perturbation	Wolfe *et al.* (1995), De Bodt (1997), Wolfe *et al.* (1997), Heman-Ackah *et al.* (2002), Halberstam (2004), and Ma and Yiu (2006)
Relative noise level	Hirano *et al.* (1986)
Residue signal power ratio	Plant *et al.* (1997)
Richness of high frequency harmonics	Hirano *et al.* (1986)
Shimmer in decibel	Wolfe *et al.* (1995), De Bodt (1997), Wolfe *et al.* (1997), and Halberstam (2004)
Signal-to-noise ratio	Yu *et al.* (2001) and Yu *et al.* (2007)
Signal-to-noise above 1000 Hz	Yu *et al.* (2001) and Yu *et al.* (2007)
Signal-to-noise ratio from Milenkovic	Wolfe and Martin (1997)
Signal-to-noise ratio from Qi	Qi *et al.* (1999) and Eadie and Doyle (2005)
Smoothed amplitude perturbation quotient	De Bodt (1997), Wolfe *et al.* (1997), and Halberstam (2004)
Smoothed cepstral peak prominence	Heman-Ackah *et al.* (2002), Halberstam (2004), and Eadie and Baylor (2006)
Smoothed pitch perturbation quotient	De Bodt (1997), Wolfe *et al.* (1997), Heman-Ackah *et al.* (2002), and Halberstam (2004)
Soft phonation index	De Bodt (1997)
Spectral distortion	Feijoo and Hernández (1990)
Spectral flatness of inverse filter	Prosek *et al.* (1987)
Spectral flatness of residue signal	Prosek *et al.* (1987) and Eadie and Doyle (2005)
Spectral noise level above and under 6000 Hz	Dejonckere and Wieneke (1996)
Spectral tilt	Eadie and Doyle (2005)
Spectral tilt of voiced segments	Eadie and Doyle (2005)
Standard deviation of cepstral peak prominence	Wolfe and Martin (1997)
Standard deviation of fundamental frequency	Wolfe *et al.* (1997)
Standard deviation of jitter	Kreiman *et al.* (1990) and Wolfe and Martin (1997)
Standard deviation of partial period comparison	Kreiman *et al.* (1990)
Standard deviation of period	Wolfe *et al.* (2000)
Standard deviation of relative frequency differences	Askenfelt and Hammarberg (1986)
Standard deviation of shimmer	Kreiman *et al.* (1990) and Wolfe and Martin (1997)
Standard deviation of signal-to-noise ratio from Milenkovic	Wolfe and Martin (1997)
Voice turbulence index	Halberstam (2004)

In the first subset there were 52 of the 69 acoustic measures on sustained vowels with a (weighted) correlation coefficient below 0.60. Weighted correlation coefficients ranged from 0.11 for coefficient of excess and voice turbulence index to 0.56 for amplitude perturbation quotient of residuals and harmonic-to-noise ratio from Yumoto. In this subset, there were 32 markers with a k-value of 2 or more. The SD_{res} statistics indicated heterogeneity for eight measures. For the remaining 24 acoustic correlates with $\overline{r_w} < 0.60$ and $k \geq 2$, SD_{res} statistics showed homogeneity. The second subset consisted of 17 acoustic measures with a (weighted) effect size equal to or above 0.60. In this subset of 17 measures, there were 4 markers with a k-value of 2 or more. Weighted correlation coefficients ranged from 0.62 for smoothed cepstral peak prominence to 0.75 for pitch amplitude. Statistical homogeneity testing (SD_{res}) indicated that these four $\overline{r_w}$-values were based on a set of homogeneous effect sizes, indicating that these effect sizes are consistently equal to or above 0.60 (smoothed cepstral peak prominence: $\overline{r_w}=0.62$, spectral flatness of residue signal: $\overline{r_w}=0.69$, Pearson r at autocorrelation peak: $\overline{r_w}=0.74$, and pitch amplitude: $\overline{r_w}=0.75$).

B. Continuous speech

Seven studies using continuous speech samples met the inclusion criteria of this meta-analysis. These studies were published in *Journal of Voice* (4), *Journal of Speech (Language) and Hearing Research* (1), *Journal of the Acoustical Society of America* (1), and *ORL* (1). As shown in Table V, there were 29 separate effect sizes pertaining to 26 distinct acoustic measures. Relevant information regarding the methodology of these seven reports is found in Table II. Whereas all seven studies used pathologic or dysphonic voice

TABLE IV. Summary of the meta-analytic findings for the individual acoustic measures of overall voice quality in sustained vowels. The acoustic measures are ordered according to their effect size (r or $\overline{r_w}$).

Acoustic measure	k [a]	r or $\overline{r_w}$ [b]	SD_{res} [c]
Fluctuation in fundamental frequency	1	0.00	/
Soft phonation index	1	0.01	/
Standard deviation of signal-to-noise ratio from Milenkovic	1	0.06	/
Frequency of second formant	1	0.07	/
Standard deviation of cepstral peak prominence	1	0.11	/
Voice turbulence index	2	0.11	He
Coefficient of excess	2	0.11	Ho
Frequency of third formant	1	0.14	/
Ratio of amplitudes of first and second harmonics	1	0.15	/
No. of harmonics	1	0.19	/
Fluctuation in amplitude	1	0.19	/
Richness of high frequency harmonics	1	0.19	/
Frequency of first formant	1	0.21	/
Standard deviation of percent jitter	2	0.22	Ho
Breathiness index	3	0.22	Ho
Spectral flatness of inverse filter	2	0.25	Ho
Fundamental frequency	3	0.28	Ho
Coefficient of variation of percent shimmer	1	0.28	/
Ratio of frequencies of second and first formants	1	0.32	/
Difference between frequencies of second and first formants	1	0.33	/
Coefficient of variation of amplitude	1	0.34	/
Standard deviation of period	2	0.37	Ho
Signal-to-noise ratio	3	0.38	He
Smoothed amplitude perturbation quotient	3	0.40	Ho
Residue signal power ratio	1	0.40	/
Relative noise level	1	0.40	/
Standard deviation of percent shimmer	2	0.41	Ho
Signal-to-noise ratio above 1000 Hz	3	0.42	He
Jitter factor	3	0.42	Ho
Power spectrum ratio	2	0.44	Ho
Amplitude perturbation quotient	6	0.45	Ho
Noise-to-harmonics ratio from MDVP	7	0.45	He
Shimmer in decibel	4	0.45	Ho
Absolute jitter	4	0.47	Ho
Standard deviation of fundamental frequency	2	0.47	Ho
Pitch perturbation quotient of residue signal	2	0.47	Ho
Coefficient of variation of percent jitter	1	0.48	/
Coefficient of variation of fundamental frequency	2	0.49	Ho
Percent jitter	6	0.49	Ho
Cepstral peak prominence	4	0.50	Ho
Natural logarithm of standard deviation of period	2	0.51	He
Percent shimmer	6	0.52	He
Spectral noise level above and under 6000 Hz	1	0.52	/
Relative average perturbation	5	0.52	Ho
Pitch perturbation quotient	3	0.52	Ho
Smoothed pitch perturbation quotient	4	0.53	Ho
Jitter ratio	2	0.53	Ho
Lyapunov coefficient	3	0.54	He
Phonatory fundamental frequency range	5	0.54	Ho
Harmonics-to-noise ratio from Yumoto	3	0.56	He
Amplitude perturbation quotient of residue signal	2	0.56	Ho
Mean harmonic emergence between 500 and 1500 Hz	1	0.58	/
Coefficient of variation of period	1	0.62	/
Smoothed cepstral peak prominence	3	0.63	Ho
Standard deviation of partial period comparison	1	0.67	/
Spectral flatness of residue signal	2	0.69	Ho
Partial period comparison	1	0.69	/
Jitter from Yumoto	1	0.71	/
Pearson r at autocorrelation peak	2	0.74	Ho
Normalized mean absolute period jitter	1	0.75	/
Pitch amplitude	3	0.75	Ho
Signal-to-noise ratio from Milenkovic	1	0.76	/
Cepstral peak magnitude	1	0.80	/
Cycle-to-cycle variation of waveform	1	0.83	/
Harmonics-to-noise ratio from Kojima	1	0.87	/
Normalized noise energy	1	0.88	/
Cepstrum of excitation signal	1	0.90	/
Spectral distortion	1	0.93	/
Normalized mean absolute period shimmer	1	0.93	/

[a] k=number of effect sizes available in the included literature.
[b] r=correlation coefficient (when $k=1$) and $\overline{r_w}$=mean weighted correlation coefficient (when $k>1$).
[c] SD_{res}=residual standard deviation, /=not applicable (when $k=1$), and Ho/He=homogeneous/heterogeneous r or $\overline{r_w}$ (when $k>1$).

samples, only three studies also investigated normal voices. The mean number of dysphonic voice samples was 50 (range 9–112). For the normal voices, the mean number was 7 (range 3–41). The mean number of subjects was 57 (range 12–153). All acoustic measures were extracted from recordings of continuous speech, most often from speakers reading from a text. With the exception of Askenfelt and Hammarberg (1986) and Ma and Yiu (2006), the so-called "rainbow passage" was read aloud and a portion (typically the second sentence) was extracted for further analysis.

As for auditory-perceptual evaluation, the mean number of judges employed across studies was 7 (range 2–16). In four studies the rating scale was an equal-appearing interval scale with 4, 6, 7, or 11 points. In two studies (Qi *et al.*, 1999; Eadie and Baylor, 2006), a visual analog scale was used. In another study (Eadie and Doyle, 2005), direct magnitude estimation was used. The following labels were used to designate the perceptual construct that was to be evaluated: hoarseness, G (for grade), overall severity, and overall voice quality. Estimates of reliability of listener judgments included two types of statistics: Pearson's product-moment correlation coefficient and Cronbach's coefficient alpha. To evaluate intrajudge reliability, Pearson's r-values were uniformly reported. Where a range of r-values was given (Askenfelt and Hammarberg, 1986; Qi *et al.*, 1999; Eadie and Baylor, 2006), the lowest r-value was chosen to calculate a weighted average of intrajudge correlation across reports (Table II). The intrajudge $\overline{r_w}$ was 0.81, which is indicative of homogeneous intrajudge reliability. It appears that listeners were generally consistent in their perceptual evaluations of continuous speech. Regarding interjudge reliability, only three studies provided a Pearson's r-value. Meta-analysis, again using the lowest r-value of the reported range, resulted in an interjudge $\overline{r_w}$ of 0.84, i.e., homogeneous interjudge reliability. This was corroborated by the three studies that used Cronbach's α, since they all mentioned an α-value of 0.97. Concerning the acoustic measures, Table V provides an overview of the determinants that were used to

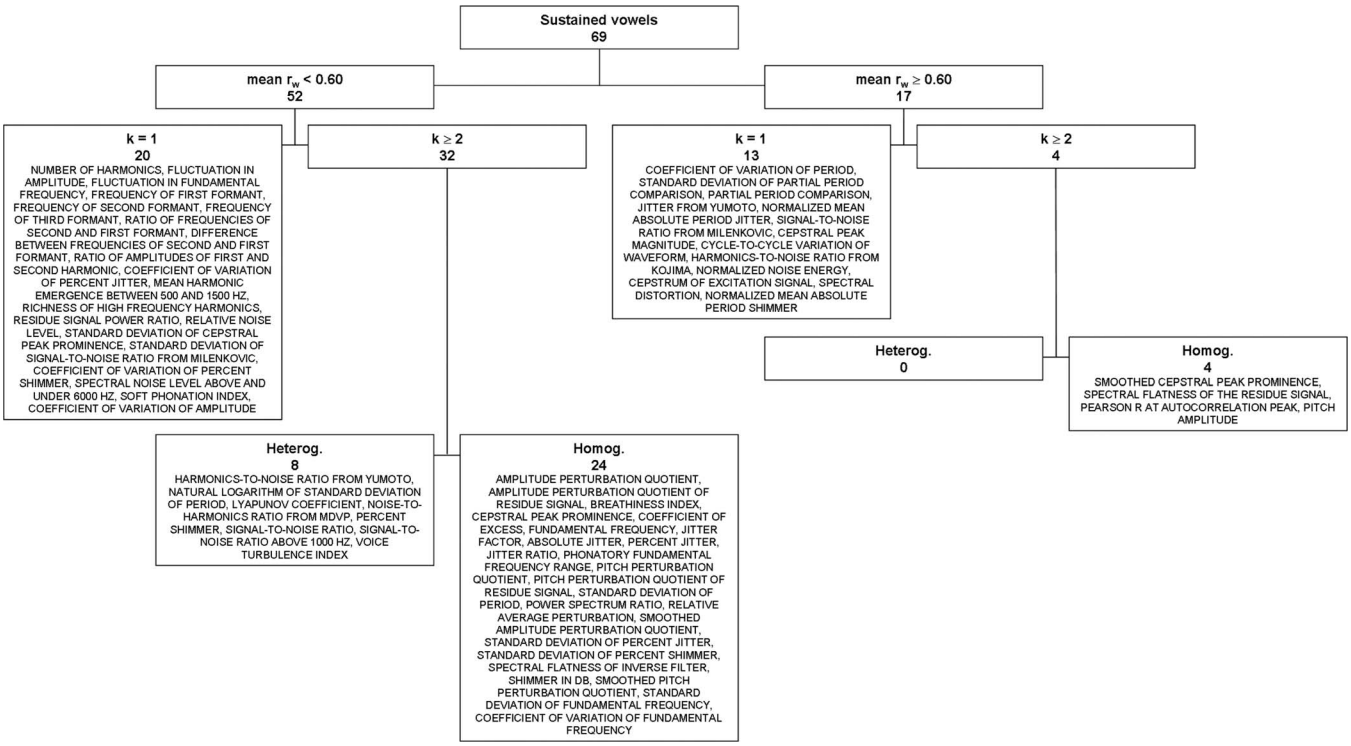


FIG. 1. Diagram illustrating the organization of the meta-analysis for acoustic measures on sustained vowels. The second line in every box contains the number of acoustic measures.

gauge overall voice quality. As was the case for sustained vowel studies, there were considerable differences between studies' recording equipment and settings.

1. Meta-analysis on correlation coefficients

The results of the meta-analysis on continuous speech data are summarized in Table V and Fig. 2. For 23 of the 26 (88%) acoustic measures cited, there was only 1 effect size available. For the remaining three acoustics determinants (cepstral peak prominence, smoothed cepstral peak prominence, and signal-to-noise ratio from Qi) there were two effect sizes ($k=2$). For these three markers, a $\overline{r_w}$ was calculated with Meta-Analysis Programs version 5.3. The organization of the meta-analysis on acoustic measures on continuous speech is illustrated in Fig. 2.

As in our meta-analysis on sustained vowel data, a correlation coefficient of 0.60 was chosen as the threshold to distinguish between marked and weak acoustic measures. In the first subset of 16 acoustic measures with a (weighted) effect size below 0.60, k was always equal to 1, and therefore no meta-analysis was performed. In the second subset consisting of ten acoustic measures with a (weighted) effect size equal to or above 0.60, there were three markers with $k=2$: signal-to-noise ratio from Qi, cepstral peak prominence, and smoothed cepstral peak prominence. Meta-analysis for these three measures yielded $\overline{r_w}$-values of 0.69, 0.88, and 0.88, respectively. Furthermore, SD_{res} indicated that these three $\overline{r_w}$-values were based on a set of homogeneous effect sizes.

IV. DISCUSSION

The present meta-analysis assessed the relationship between acoustic measures and perceptual judgments of overall voice quality. In Buder (2000) alone, more than 100 acoustic algorithms were cited and numerous microcomputer-based software systems offering various acoustic voice quality parameters have been developed. The fact that correlations between perception of overall voice quality and acoustic measures vary substantially (Kreiman and Gerratt, 2000a) raises questions regarding the validity and usefulness of these acoustic determinants. This meta-analysis represented an attempt to synthesize the corpus of algorithms and measures, and to establish a hierarchy of acoustic markers on a statistical basis. In total, 25 study reports were included. 21 studies reported on 150 correlation coefficients for 69 acoustic measures on sustained vowels. 7 studies identified 29 correlation coefficients for 26 acoustic measures on continuous speech.

In the context of the present meta-analysis, a homogeneous $\overline{r_w}$ exceeding 0.60 was judged to be a critical index. For instance, the amplitude perturbation quotient measure on sustained vowels was cited in five studies with 0.41, 0.54, 0.63, 0.50, 0.41, and 0.71 as coefficients of correlation. The single $r=0.71$ value, in particular, (Halberstam, 2004) seems to identify amplitude perturbation quotient as a valid acoustic marker for overall voice quality of sustained vowels. However, the r-values from the other studies are less persuasive and the meta-analysis resulted in a smaller homogeneous $\overline{r_w}$ of 0.45. In contrast to the amplitude perturbation quotient example wherein the meta-analysis resulted in a relatively weak $\overline{r_w}$ of 0.45, the meta-analysis outcome of

TABLE V. Summary of the meta-analytic findings for the individual acoustic measures of overall voice quality in continuous speech. The acoustic measures are ordered according to their effect size (r or $\overline{r_w}$).

Acoustic measure	k [a]	r or $\overline{r_w}$ [b]	SD_{res} [c]
Perturbation magnitude	1	0.01	/
Maximum intensity in voice range profile	1	0.02	/
Lowest fundamental frequency in voice range profile	1	0.09	/
Noise-to-harmonics ratio from MDVP	1	0.13	/
Fundamental frequency	1	0.18	/
Frequency-domain harmonics-to-noise ratio	1	0.26	/
Spectral flatness of residue signal	1	0.26	/
Spectral tilt of voiced segments	1	0.33	/
Highest fundamental frequency in voice range profile	1	0.34	/
Intensity range in voice range profile	1	0.35	/
Fundamental frequency range in voice range profile	1	0.37	/
Minimum intensity in voice range profile	1	0.38	/
Area of voice range profile	1	0.43	/
Spectral tilt	1	0.47	/
Pitch amplitude	1	0.58	/
Perturbation magnitude mean	1	0.59	/
Perturbation factor	1	0.62	/
Percent shimmer	1	0.62	/
Signal-to-noise ratio from Qi	2	0.69	He
Directional perturbation factor	1	0.71	/
Standard deviation of relative frequency differences	1	0.71	/
Compression of relative frequency differences	1	0.73	/
Peakedness of relative frequency differences	1	0.73	/
Relative average perturbation	1	0.75	/
Cepstral peak prominence	2	0.88	Ho
Smoothed cepstral peak prominence	2	0.88	Ho

[a] k=number of effect sizes available in the included literature.
[b] r=correlation coefficient (when $k=1$) and $\overline{r_w}$=mean weighted correlation coefficient (when $k>1$).
[c] SD_{res}=residual standard deviation, /=not applicable (when $k=1$), and Ho/He=homogeneous/heterogeneous r or $\overline{r_w}$ (when $k>1$).

studies related to smoothed cepstral peak prominence seems to suggest a much stronger association. For instance, although Halberstam's (2004) r-value of 0.55 for smoothed cepstral peak prominence does not provide strong support for smoothed cepstral peak prominence as a valid measure of overall voice quality; combining this result with the Heman-Ackah *et al.* (2002) and the Eadie and Baylor (2006) results of $r=0.80$ and $r=0.82$, respectively, the final $\overline{r_w}$ is 0.63, which supports smoothed cepstral peak prominence as a promising acoustic marker of overall voice quality. Based on the meta-analysis of sustained vowel studies, four measures satisfied the requirement of a homogeneous $\overline{r_w} \geq 0.60$: (1) Pearson r at autocorrelation peak, (2) pitch amplitude, (3) spectral flatness of residue signal, and (4) smoothed cepstral peak prominence. For continuous speech, three measures satisfied the criterion: (1) signal-to-noise ratio from Qi, (2) cepstral peak prominence, and (3) smoothed cepstral peak prominence. Consequently, these six measures are considered to be the most promising measures for the acoustic measurement of overall voice quality, as compared to the remaining 81 measures included in the original meta-analysis. The results of these six measures will be discussed in the next paragraphs.

The first of these six measures is Pearson r at autocorrelation peak. To obtain this measure, correlations are calculated between the voice signal and delayed versions of the same signal (i.e., autocorrelation) at time lags between the minimally and maximally expected fundamental periods. The Pearson moment-product correlation coefficient is computed at the highest peak of this autocorrelation function (i.e., the correlogram with "delay" or "time lag" on the abscissa and "correlation" on the ordinate). The rationale behind this measure is that more periodic voice signals display more prominent autocorrelation peaks, and vice versa. A perfectly periodic signal reveals a Pearson r at autocorrelation peak of 1.0, and the more the signal deviates from perfect periodicity, the more this correlation decreases (Hillenbrand and Houde, 1996). The correlation between overall voice quality and this measure of the autocorrelation function on the sound waveform has been investigated by Wolfe *et al.* (2000) in both male and female voices separately, yielding $k=2$ and $\overline{r_w}=0.74$. This result requires confirmation by other independent investigators to permit generalization. Although Hillenbrand and Houde (1996) indicated a correlation of 0.84 between breathiness ratings and Pearson r at autocorrelation peak for both sustained vowels and continuous speech, and

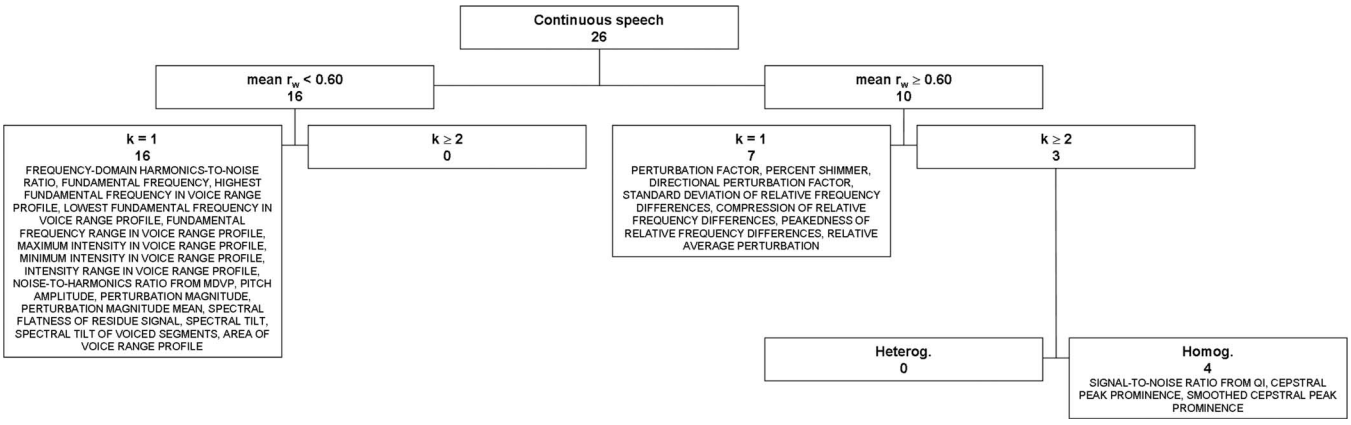


FIG. 2. Diagram illustrating the organization of the meta-analysis for acoustic measures on continuous speech. The second line in every box contains the number of acoustic measures.

concluded that Pearson r at autocorrelation peak is an accurate marker of breathiness, further corroboration of its concurrent validity is also needed.

The second measure is pitch amplitude. To acquire this measure, the radiated voice signal is first inverse filtered via a linear predictive coding algorithm. The result of this inverse filtering is a residue signal, i.e., a series of impulses theoretically showing the moment of vocal tract acoustic excitation provided by glottal closure (permitting investigation of the signal provided by the laryngeal source instead of the entire vocal tract). Second, the autocorrelation function of this residue signal is calculated. Pitch amplitude is the amplitude of the maximum correlation (i.e., traditionally corresponding with the pitch) in the correlogram and consequently is considered to be a measure of the strength of voice periodicity (Prosek *et al.*, 1987). Plant *et al.* (1997) investigated the correlation between this measure and overall voice quality, and Prosek *et al.* (1987) used pitch amplitude as a marker of both disorder severity and hoarseness in sustained vowels. These two independent studies resulted in $k=3$ and $\overline{r_w}=0.75$. Although Eadie and Doyle (2005) reported an r-value for pitch amplitude of only 0.58 when applied on continuous speech, further support for the value of measures based on inverse filtering is provided by Parsa and Jamieson (2001), who concluded that such measures are superior to perturbation measures for both continuous speech and sustained vowels. In part, Parsa and Jamieson (2001) arrived at their conclusion based on measures of diagnostic accuracy, which included the area under the ROC curve. In the case for pitch amplitude, the area under the ROC curve for sustained vowels was 0.977 (perfect diagnostic accuracy=1.00), and the rate of correct classification between normal and pathologic voices was 93.0%. For continuous speech there was an area under ROC curve of 0.953 and a correct classification rate of 88.9%. Parsa and Jamieson (2001) stated that pitch amplitude provided the best classification accuracy among all measures extracted from continuous speech samples.

Based on the results of the meta-analysis, the third acoustic measure, which produced respectable raw correlation results, was spectral flatness of residue signal. This measure also generates the residue signal as an output of the inverse filter. The spectrum is then derived from the residue signal, and finally the distribution of the frequencies in the spectrum is computed. The flatter the spectral distribution of the residue signal, the more the harmonics are considered to be masked by noise (Prosek *et al.*, 1987). Prosek *et al.* (1987) correlated spectral flatness of residue signal with both disorder severity and hoarseness separately ($k=2$) in sustained vowels. Our meta-analysis of these two correlation coefficients (derived from one study) yielded $\overline{r_w}=0.69$. Although there is no confirmation from other independently generated correlations, Parsa and Jamieson (2001), who used discriminant analyses, supported the utility of spectral flatness of residue signal as a valid discriminator between normal and pathological voices. For sustained vowels, spectral flatness of residue signal showed the largest area under ROC curve (0.996) and had the highest classification accuracy (96.5% correct). For continuous speech, the area under the ROC and the discrimination accuracy were 0.928% and 85.8%, respectively. Furthermore, Parsa and Jamieson (2001) concluded that more commonly used measures (as jitter, shimmer, and noise-to-harmonics ratio) did not perform as well as measures based on linear prediction modeling and inverse filtering.

The concurrent validity of the fourth measure, signal-to-noise ratio from Qi, was investigated in continuous speech only. This measure uses linear predictive coding and inverse filtering for the decomposition of speech samples into signal (i.e., waveform of the original signal) and noise (i.e., waveform of the signal with a typically random Gaussian distribution after removal of resonance-based and voice-based patterns). The ratio between the average root-mean-square amplitudes of the signal and the noise components can then be computed to quantify the acoustic properties of disordered voices (Qi *et al.*, 1999). Combining the independent results of Qi *et al.* (1999) and Eadie and Doyle (2005) leads to a homogeneous $\overline{r_w}=0.69$ ($k=2$), which is promising. This measure was also examined in the studies of Parsa and Jamieson (2001). Although not as robust as pitch amplitude and spectral flatness of residue signal, signal-to-noise ratio from Qi demonstrated acceptable diagnostic precision (distinguishing normophonic from dysphonic individuals) in both sustained vowels (area under ROC curve: 0.945; classification rate: 81.6%) and continuous speech (area under ROC curve: 0.903; classification rate: 79.6%). In summary, measures and algorithms based on inverse filtering and linear prediction modeling appear to be very promising and useful in clinical settings, where patients present with heterogeneous voice qualities and severities.

The meta-analysis outcome for the fifth and sixth measures, the cepstral markers of cepstral peak prominence and smoothed cepstral peak prominence, can be summarized as follows. To obtain these two measures, one constructs a cepstrum (i.e., a log power spectrum of a log power spectrum, resulting in a graph with "quefrency" on the abscissa and "cepstral magnitude" on the ordinate). The highest cepstral peak is identified between the minimally and maximally expected fundamental period, and a linear regression line is drawn, which relates quefrency to cepstral magnitude. The difference in amplitude between this cepstral peak and the corresponding value on the linear regression line exactly below the peak determines the cepstral peak prominence. Averaging (i.e., smoothing) of the cepstrum across time and across quefrency results in a smoothed cepstrum, and the difference between the highest peak and the corresponding value on the regression line in the smoothed cepstrum is called the smoothed cepstral peak prominence. The rationale behind these measures is that the more periodic a voice signal is, the more it displays a well-defined harmonic configuration in the spectrum, and, consequently, the more the cepstral peak will be prominent (Hillenbrand and Houde, 1996). For cepstral peak prominence on sustained vowels, meta-analysis of the Wolfe and Martin (1997), the Wolfe *et al.* (2000), and the Halberstam (2004) results yields a homogeneous $\overline{r_w}$ of 0.50 ($k=3$). On continuous speech, however, meta-analysis on the findings of Halberstam (2004) and Eadie and Baylor (2006) results in $\overline{r_w}=0.88$ ($k=2$). Heman-Ackah *et al.* (2002), Halberstam (2004), and Eadie and

Doyle (2005) investigated smoothed cepstral peak prominence applied on sustained vowels. Meta-analysis of these three independent studies results in $\overline{r_w}=0.63$ ($k=3$). Furthermore, Heman-Ackah *et al.* (2002) and Halberstam (2004) also provided correlation coefficients for continuous speech, which results in a homogeneous $\overline{r_w}=0.88$ ($k=2$) after meta-analysis. In summary, on the basis of this meta-analysis, the two cepstral measures, and smoothed cepstral peak prominence, in particular, can be viewed as potentially the most accurate acoustic algorithms or single correlates of overall voice quality. Additional evidence for the validity of cepstral measures can be found in Hillenbrand and Houde (1996), who found that cepstral peak prominence was among the most robust correlates of breathiness in sustained vowels as well as in continuous speech. Other studies confirming this conclusion were conducted by de Krom (1993), who stated that cepstrum-based harmonics-to-noise ratio was a strong marker of both roughness and breathiness in sustained vowels. Dejonckere and Wieneke (1996) found a correlation of 0.80 between overall severity of hoarseness and the amplitude of highest rahmonic (also known as cepstral peak magnitude). The magnitude of this correlation far exceeded the correlation of the other acoustic measures in their study (jitter ratio, relative noise level above 6 kHz, and mean harmonic emergence between 0.5 and 1.5 kHz). In a later study using factor analysis, Dejonckere (1998) reported that cepstral peak magnitude is negatively affected by irregularity in vocal fold vibration as well as by excessive glottal air leakage, bolstering the assertion that cepstral peak magnitude is sensitive to aspects that potentially contribute to overall dysphonia severity. Heman-Ackah *et al.* (2003) investigated the diagnostic validity of smoothed cepstral peak prominence on sustained vowels and continuous speech and of amplitude perturbation quotient, percent jitter, noise-to-harmonics ratio from MDVP, relative average perturbation, and smoothed pitch perturbation quotient on sustained vowels only. They concluded that the smoothed cepstral peak prominence measures are good correlates of dysphonia and that, on average, smoothed cepstral peak prominence on continuous speech performed better on measures of diagnostic precision such as sensitivity, specificity, positive predictive value, and negative predictive value, as compared to traditional time-based measures of perturbation. They concluded that smoothed cepstral peak prominence "are reliable measures that should become routine in objective voice analysis" (p. 332). Finally, Awan and Roy (2006) conducted a study in which they used a cepstral measure they called expected cepstral peak prominence in a multiple regression procedure. This measure actually is the ratio of the cepstral peak prominence to the expected amplitude of the cepstral peak based on linear regression. It is very similar to the cepstral peak prominence measure described by Hillenbrand and Houde (1996). Awan and Roy (2006) indicated that expected cepstral peak prominence "may be the most significant component" (p. 44) contributing to their four-factor model for measuring dysphonia severity. Collectively, measures derived from the cepstrum (such as cepstral peak prominence and smoothed cepstral peak prominence) can be used in sustained vowel as well as continuous speech samples because they do not rely on accurate fundamental period detection (Hillenbrand and Houde, 1996; Heman-Ackah *et al.*, 2003), and they can be easily implemented in clinical settings.

From the discussion above it is apparent that four of these six most promising acoustic measures share one interesting feature, i.e., they all measure what might be called "periodicity prominence." Cepstral peak prominence and smoothed cepstral peak prominence are two very similar measures initially introduced for pitch detection in speech signals via the cepstral method, and pitch amplitude and Pearson *r* at autocorrelation peak are two similar measures frequently applied for pitch determination via the autocorrelation function. Both the quefrency of the first rahmonic and the lag time of the highest autocorrelation peak within a specific analysis window correspond with the fundamental frequency; and, the height of these cepstral and autocorrelation peaks is related to the prominence of the fundamental frequency (i.e., periodicity) in the voice signal (Hillenbrand *et al.*, 1994). Based on the outcome of the present meta-analysis, it thus can be assumed that overall voice quality and dysphonia severity are mainly determined by periodicity dominance, and that factors attenuating periodicity of the voice signal also contribute to the perception of increased dysphonia. Furthermore, it is important to note that there are many other measures of periodicity prominence available to clinicians, such as the ubiquitous pitch and amplitude perturbation measures often generated automatically by most commercially available signal processing software. However, the important advantage of these four measures over the commonly used voice perturbation measures resides in their methods to assess periodicity, namely, they do not demand cycle boundary identification for fundamental frequency detection in the time-domain.

In addition to these six $k>1$ measures with $\overline{r_w}\geq 0.60$, there were many $k=1$ measures. However, because a high correlation in one study can be offset by a low correlation in another study (and vice versa), caution is warranted when interpreting the outcome of a solitary *r*, and this applies to all acoustic measures with $k=1$ (e.g., $r=0.93$ for normalized mean absolute period shimmer on sustained vowels or $r=0.26$ for frequency-domain harmonics-to-noise ratio on continuous speech). Without further confirmation, replication, or evidence rejecting the presented *r*-values, it is difficult to place these results in context, and impossible to draw firm conclusions regarding these $k=1$ measures at this point in time.

Interestingly, the measures that were investigated most often (with $k\geq 5$) were the perturbation measures (amplitude perturbation quotient, percent jitter, percent shimmer, and relative average perturbation), the noise measure noise-to-harmonics ratio from MDVP, and voice range profile measure phonatory fundamental frequency range (see Table V). Except for the phonatory fundamental frequency range, these time-domain perturbation measures also appeared to be most frequently used in voice clinics (De Bodt, 1997). Percent jitter is a measure of fundamental frequency or period perturbation. It measures the mean difference in fundamental frequency of adjacent periods relative to the mean fundamental frequency of all periods in the voice recording (Buder,

2000). Relative average perturbation is also a measure of fundamental frequency perturbation. This measure is similar to percent jitter, but uses a moving three-point smoothing and normalization to average the period before computing the mean deviation in period relative to the mean period of all periods (Buder, 2000). Percent shimmer, another measure similar to percent jitter, measures amplitude perturbations by computing the mean deviation in amplitude between adjacent cycles relative to the mean amplitude of all cycles (Buder, 2000). Amplitude perturbation quotient is another amplitude perturbation measure, but instead of working with adjacent cycles as percent shimmer, it first averages the amplitude of a moving number (i.e., an odd integer greater than 1) of successive cycles before calculating the mean deviation in amplitude between cycle groups relative to the mean amplitude of all cycles (Buder, 2000). These perturbation measures are traditionally linked to the measurement of irregular voice fold vibrations. The noise-to-harmonics ratio from MDVP is a spectral measure that computes the ratio of the between-harmonic spectral magnitudes in the range from 1500 to 4500 Hz to the harmonic spectral magnitudes in the range from 70 to 4500 Hz (Buder, 2000). This measure is classically associated with measurements of additive noise at the level of the glottis. The phonatory fundamental frequency range in the voice range profile is one of the measures of the dispersion of the fundamental frequency and consists of subtracting the lowest from the highest possible fundamental frequency (Buder, 2000). According to De Bodt (1997), who reviewed literature between 1991 and 1995, these are the most frequently mentioned measures in voice literature (except for F_0 and amplitude measures). Yet, on sustained vowels, these measures did not yield $\overline{r_w} \geq 0.60$. Regarding jitter, meta-analysis yielded homogeneous $\overline{r_w}$ of 0.47, 0.49, 0.52, and 0.52 for absolute jitter, percent jitter, relative average perturbation, and pitch perturbation quotient, respectively. Absolute jitter is the mean of the differences between the period and the fundamental frequency of adjacent cycles (Buder, 2000). Pitch perturbation quotient is the same as relative average perturbation, but with a smoothing factor of 5 cycles (Buder, 2000). Similarly, the meta-analysis for shimmer resulted in a homogeneous $\overline{r_w}$ of 0.45 for shimmer in decibel and amplitude perturbation quotient, and a heterogeneous $\overline{r_w}$ of 0.52 for percent shimmer. Regarding noise-to-harmonics ratio from MDVP, the measure most frequently encountered, a heterogeneous $\overline{r_w}$ of 0.45 was found. On continuous speech, there was a solitary correlation of 0.62 and 0.75 for percent shimmer and relative average perturbation, and 0.37 and 0.13 for phonatory fundamental frequency range and noise-to-harmonics ratio from MDVP, respectively. Measures related to the voice range profile yielded a $\overline{r_w}$ of maximally 0.43. In general, the results of this meta-analysis confirm the apparent inferiority of perturbation measures as compared to other measures that do not depend on accurate identification of cycle boundaries. This conclusion supports the findings of Parsa and Jamieson (2001), and is confirmed by Kreiman and Gerratt (2005), who concluded that "the associations between jitter, shimmer, and perceived voice quality are not sufficiently explanatory to justify continued reliance on jitter and shimmer as indices of voice quality" (p. 2209). As mentioned previously, F_0 and amplitude perturbation measures are especially susceptible for the influence of type of microphone and microphone location relative to the sound source, type of hardware, processing algorithms, measurement algorithms, and software settings such as sampling rate and fundamental period extraction. Furthermore, F_0 and amplitude perturbation measures are not sensitive to differences in glottal waveform shape and additive glottal noise, and appear only reliable in nearly periodic voice signals (Titze, 1995; Parsa and Jamieson, 2001).

This meta-analysis, combined with previous studies, seems to confirm that measures that do not rely on the extraction of the fundamental period in their calculation such as smoothed cepstral peak prominence, Pearson r at autocorrelation peak, and pitch amplitude produce stronger relationships with perceptual judgments of overall severity of dysphonia in sustained vowels as well as continuous speech, and deserve further attention in clinical circles (Hillenbrand and Houde, 1996; Parsa and Jamieson, 2001).

A. Caveats and limitations

There are several limitations regarding the present meta-analysis that not only restrict the generalizability of the findings, but also provide a direction for future research. It is important to acknowledge that current acoustic measures might not be sensitive measures of perceived voice quality because of limitations of their algorithms and the theoretical models on which they are based. First, this meta-analysis concentrated on the relationship between acoustic markers and overall voice quality. Additional meta-analytic research is needed to address the relationship between acoustic measures and specific vocal quality attributes, such as breathiness and roughness. Meta-analytic techniques may improve the resolution regarding which acoustic measures best track these specific voice qualities.

Second, the present meta-analysis was restricted to reports and findings based on correlation coefficients. Beyond the 69 acoustic measures on sustained vowels and the 26 measures on continuous speech, other acoustic measures have been discussed in literature. But because correlation coefficients were not available for these measures, the value of these markers in voice quality measurement and their relative validity in comparison to the aforementioned markers remains unclear. Future meta-analyses should potentially explore other effect size measures, aside from the correlation coefficient, to investigate the validity of these acoustic markers.

Third, overall voice quality can be investigated with measures other than acoustic measures. For example, certain aerodynamic measures could also be worth exploring within this context, and thus meta-analysis investigating the association between aerodynamic measures and perceptual voice quality measurement is recommended.

Fourth, the interpretation of the findings of the present meta-analysis is complicated by variability related to different data acquisition systems. While the influence of factors such as microphone type and placement, environmental noise, software, etc., on the outcomes of perturbation mea-

sures has already been investigated, the impact of these factors on other measures such as cepstral peak prominence and pitch amplitude remains unclear. Additional exploration of the impact of data acquisition systems and environments on the outcomes of these measures is warranted.

Fifth, the relationship between the auditory-perceptual rating and the acoustic measurement of overall voice quality relies greatly on the rationale and algorithm underlying the acoustic measure. However, as previously discussed, unreliability of listener ratings introduces perceptual noise and consequently tends to handicap the acoustic (or other) measurement of voice quality. While suggestions to improve rater reliability exist (e.g., Kreiman and Gerratt, 2000b; Eadie and Doyle, 2002; Bele, 2005; Eadie and Baylor, 2006; Yiu *et al.*, 2007; Kreiman *et al.*, 2007) few studies have estimated the true (absolute) impact of listener unreliability on the correlation between perception and acoustic measures. Furthermore, there is no universal standard distinguishing an acceptable from an unacceptable reliability estimate. Future research should address the criteria used to determine what precisely constitutes an acceptable level of listener reliability, and the impact of such criteria on the validation of acoustic voice quality measures.

Sixth, an important caveat is related to the low number of correlation coefficients (i.e., the k statistic) available for many of the acoustic measures in this meta-analysis. Because of $k=1$, no weighted average correlation coefficient could be calculated for 33 of the measures on sustained vowels (47.8%) and for 23 of the measures on continuous speech (88.5%). However, as long as the extant literature lacks corroborating evidence from multiple, independent correlational studies, no firm conclusions can be made regarding this very diverse and idiosyncratic set of $k=1$ acoustic measures. Although impressive correlations with dysphonia severity have been reported for some of these $k=1$ measures in sustained vowels (e.g., $r=0.88$ for normalized noise energy) as well as continuous speech (e.g., $r=0.75$ for relative average perturbation), equally poor correlations have been reported for others, such as $r=0.01$ for soft phonation index for sustained vowels, and $r=0.01$ for perturbation magnitude for connected speech. There were also several $k>1$ measures with restricted interpretative value, because although multiple r-values were reported, they were derived from the same study report. For instance, the two correlations on which the meta-analyses on Pearson r at autocorrelation peak (Wolfe *et al.*, 2000) and spectral flatness of residue signal (Prosek *et al.*, 1987) were based actually originated from the single studies of Wolfe *et al.* (2000) and Prosek *et al.* (1987), respectively. Furthermore, two of the three effect sizes in the meta-analysis on pitch amplitude also originated from the single study of Prosek *et al.* (1987). Because these are not replications *per se*, the generalizability of the meta-analytic evidence for these three measures is also limited. Therefore, like the $k=1$ scenario, further research/replication is also needed to corroborate the performance of these three acoustic measures.

V. CONCLUSIONS

The above-stated limitations notwithstanding, measures for which the meta-analysis resulted in a homogeneous $\overline{r_w}$ of at least 0.60, are Pearson r at autocorrelation peak, pitch amplitude, spectral flatness of residue signal, and smoothed cepstral peak prominence on sustained vowels; and signal-to-noise ratio from Qi, cepstral peak prominence, and smoothed cepstral peak prominence on continuous speech. However, only the smoothed cepstral peak prominence withstanded all criteria demanded by this meta-analytic approach: multiple r-values, derived from multiple study reports, leading to homogeneous $\overline{r_w} \geq 0.60$ in both sustained vowels and continuous speech. This cepstral metric thus can be regarded as the most promising and perhaps robust acoustic measure of dysphonia severity. Tables IV and V present a hierarchy of the numerous outcomes of acoustic markers measuring overall voice quality, but the reader is directed to the height of r or $\overline{r_w}$ as a quantity-based overview of the domain of acoustic voice quality measurement. Furthermore, the tables show the relative position of a given acoustic measure according to its concurrent validity as a measure of overall voice quality. In this regard, the present meta-analysis was able to effectively distil an extremely large number of potential acoustic measures to a subset of strong independent variables. This should be particularly informative for voice practitioners in clinical settings who are faced with software packages that automatically generate a daunting number of acoustic measures ostensibly aimed to quantify dysphonia severity and track voice change following intervention. The present meta-analysis confirmed that not all acoustic measures are created equal with respect to these clinical goals.

ACKNOWLEDGMENTS

The assistance by Jan Deman (Medical Library, Sint-Jan General Hospital, Bruges, Belgium) for library work and article retrieval is greatly appreciated. The authors also would like to credit the associate editor and the three anonymous reviewers for the numerous valuable comments on earlier versions of this manuscript.

Askenfelt, A. G., and Hammarberg, B. (**1986**). "Speech waveform perturbation analysis: A perceptual-acoustical comparison of seven measures," J. Speech Hear. Res. **29**, 50–64.

Awan, S. N., and Roy, N. (**2006**). "Toward the development of an objective index of dysphonia severity: A four-factor acoustic model," Clin. Linguist. Phonetics **20**, 35–49.

Bele, I. V. (**2005**). "Reliability in perceptual analysis of voice quality," J. Voice **19**, 555–573.

Buder, E. H. (**2000**). "Acoustic analysis of voice quality: A tabulation of algorithms 1902–1990," in *Voice Quality Measurement*, edited by R. D. Kent and M. J. Ball (Singular, San Diego, CA), pp. 119–244.

De Bodt, M. (**1997**). "A framework of voice assessment: The relation between subjective and objective parameters in the judgment of normal and pathological voice," Ph.D. thesis, University of Antwerp, Antwerp, Belgium.

de Krom, G. (**1993**). "A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals," J. Speech Hear. Res. **36**, 254–266.

Dejonckere, P. H. (**1998**). "Cepstral voice analysis: Link with perception and stroboscopy," Rev. Laryngol. Otol. Rhinol. (Bord) **119**, 245–246.

Dejonckere, P. H., Remacle, M., Fresnel-Elbaz, E., Woisard, V., Crevier-Buchman, L., and Millet, B. (**1996**). "Differentiated perceptual evaluation of pathological voice quality: Reliability and correlations with acoustic

measurements," Rev. Laryngol. Otol. Rhinol. (Bord) **117**, 219–224.
Dejonckere, P. H., and Wieneke, G. H. (**1996**). "Cepstra of normal and pathological voices: Correlation with acoustic, aerodynamic and perceptual data," in *Advances in Clinical Phonetics*, Studies in Speech Pathology and Clinical Linguistics Vol. **6**, edited by M. J. Ball and M. Duckworth (John Benjamins, Amsterdam), pp. 217–227.
Eadie, T. L., and Baylor, C. R. (**2006**). "The effect of perceptual training on inexperienced listeners' judgments of dysphonic voice," J. Voice **20**, 527–544.
Eadie, T. L., and Doyle, P. C. (**2002**). "Direct magnitude estimation and interval scaling of pleasantness and severity in dysphonic and normal speakers," J. Acoust. Soc. Am. **112**, 3014–3021.
Eadie, T. L., and Doyle, P. C. (**2005**). "Classification of dysphonic voice: Acoustic and auditory-perceptual measures," J. Voice **19**, 1–14.
Feijoo, S., and Hernández, C. (**1990**). "Short-term stability measures for the evaluation of vocal quality," J. Speech Hear. Res. **33**, 324–334.
Franzblau, A. N. (**1958**). *A Primer of Statistics for Non-Statisticians* (Harcourt, Brace & Company, New York).
Frey, L. R., Botan, C. H., Friedman, P. G., and Kreps, G. L. (**1991**). *Investigating Communication: An Introduction to Research Methods* (Prentice-Hall, Englewood Cliffs, NJ).
Gorham-Rowan, M. M., and Laures-Gore, J. (**2006**). "Acoustic-perceptual correlates of voice quality in elderly men and women," J. Commun. Disord. **39**, 171–184.
Halberstam, B. (**2004**). "Acoustic and perceptual parameters relating to connected speech are more reliable measures of hoarseness than parameters relating to sustained vowels," ORL **66**, 70–73.
Heman-Ackah, Y. D., Heuer, R. J., Michael, D. D., Ostrowski, R., Horman, M., Baroody, M. M., Hillenbrand, J., and Sataloff, R. T. (**2003**). "Cepstral peak prominence: A more reliable measure of dysphonia," Ann. Otol. Rhinol. Laryngol. **112**, 324–333.
Heman-Ackah, Y. D., Michael, D. D., and Goding, G. S. (**2002**). "The relationship between cepstral peak prominence and selected parameters of dysphonia," J. Voice **16**, 20–27.
Hillenbrand, J., Cleveland, R. A., and Erickson, R. L. (**1994**). "Acoustic correlates of breathy vocal quality," J. Speech Hear. Res. **37**, 769–778.
Hillenbrand, J., and Houde, R. A. (**1996**). "Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech," J. Speech Hear. Res. **39**, 311–321.
Hirano, M., Hibi, S., Terasawa, R., and Masako, F. (**1986**). "Relationship between aerodynamic, vibratory, acoustic and psychoacoustic correlates in dysphonia," J. Phonetics **14**, 445–456.
Hunter, J. E., Schmidt, F. L., and Jackson, G. B. (**1982**). *Meta-Analysis, Cumulating Research Findings Across Studies* (Sage, Beverly Hills, CA).
Kojima, H., Gould, W. J., Lambiase, A., and Isshiki, N. (**1980**). "Computer analysis of hoarseness," Acta Oto-Laryngol. **89**, 547–554.
Kreiman, J., and Gerratt, B. (**2000a**). "Measuring vocal quality," in *Voice Quality Measurement*, edited by R. D. Kent and M. J. Ball (Singular, San Diego, CA), pp. 73–101.
Kreiman, J., and Gerratt, B. (**2005**). "Perception of aperiodicity in pathological voice," J. Acoust. Soc. Am. **117**, 2201–2211.
Kreiman, J., and Gerratt, B. R. (**2000b**). "Sources of listener disagreement in voice quality assessment," J. Acoust. Soc. Am. **108**, 1867–1876.
Kreiman, J., Gerratt, B. R., and Ito, M. (**2007**). "When and why listeners disagree in voice quality assessment tasks," J. Acoust. Soc. Am. **122**, 2354–2364.
Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., and Berke, G. S. (**1993**). "Perceptual evaluation of voice quality: Review, tutorial, and a framework for research," J. Speech Hear. Res. **36**, 21–40.
Kreiman, J., Gerratt, B. R., and Precoda, K. (**1990**). "Listener experience and perception of voice quality," J. Speech Hear. Res. **33**, 103–115.
Lipsey, M. W., and Wilson, D. B. (**2001**). *Practical Meta-Analysis* (Sage, Thousand Oaks, CA).
Ma, E., and Yiu, E. (**2006**). "Multiparametric evaluation of dysphonic severity," J. Voice **20**, 380–390.
Parsa, V., and Jamieson, D. G. (**2001**). "Acoustic discrimination of pathological voice: Sustained vowels versus continuous speech," J. Speech Lang. Hear. Res. **44**, 327–339.
Plant, R. L., Hillel, A. D., and Waugh, P. F. (**1997**). "Analysis of voice changes after thyroplasty using linear predictive coding," Laryngoscope **107**, 703–709.
Prosek, R. A., Montgomery, A. A., Walden, E., and Hawkins, D. B. (**1987**). "An evaluation of residue features as correlates of voice disorders," J. Commun. Disord. **20**, 105–117.
Qi, Y., Hillman, R. E., and Milstein, C. (**1999**). "The estimation of signal-to-noise ratio in continuous speech for disordered voices," J. Acoust. Soc. Am. **105**, 2532–2535.
Titze, I. R. (**1995**). *Workshop on Acoustic Voice Analysis: Summary Statement* (National Center for Voice and Speech, Iowa City, IA).
Wolfe, V., Fitch, J., and Cornell, R. (**1995**). "Acoustic prediction of severity in commonly occurring voice problems," J. Speech Hear. Res. **38**, 273–279.
Wolfe, V., Fitch, J., and Martin, D. (**1997**). "Acoustic measures of dysphonic severity across and within voice types," Folia Phoniatr Logop **49**, 292–299.
Wolfe, V., and Martin, D. (**1997**). "Acoustic correlates of dysphonia: Type and severity," J. Commun. Disord. **30**, 403–416.
Wolfe, V. I., Martin, D. P., and Palmer, C. I. (**2000**). "Perception of dysphonic voice quality by naïve listeners," J. Speech Lang. Hear. Res. **43**, 697–705.
Wolfe, V. I., and Steinfatt, T. M. (**1987**). "Prediction of vocal severity within and across voice types," J. Speech Hear. Res. **30**, 230–240.
Yiu, E. M., Chan, K. M., and Mok, R. S. (**2007**). "Reliability and confidence in using a paired comparison paradigm in perceptual voice quality evaluation," Clin. Linguist. Phonetics **21**, 29–45.
Yu, P., Garrel, R., Nicollas, R., Ouaknine, M., and Giovanni, A. (**2007**). "Objective voice analysis in dysphonic patients: New data including nonlinear measurements," Folia Phoniatr Logop **59**, 20–30.
Yu, P., Ouaknine, M., Revis, J., and Giovanni, A. (**2001**). "Objective voice analysis for dysphonic patients: A multiparametric protocol including acoustic and aerodynamic measurements," J. Voice **15**, 529–542.
Yumoto, E., Sasaki, Y., and Okamura, H. (**1984**). "Harmonics-to-noise ratio and psychophysical measurement of the degree of hoarseness," J. Speech Hear. Res. **27**, 2–6.
Zraick, R. I., Wendel, K., and Smith-Olinde, L. (**2005**). "The effect of speaking task on perceptual judgment of the severity of dysphonic voice," J. Voice **19**, 574–581.

Microscopic prediction of speech recognition for listeners with normal hearing in noise using an auditory model[a)]

Tim Jürgens and Thomas Brand
Medizinische Physik, Universität Oldenburg, D-26111 Oldenburg, Germany

(Received 6 June 2008; revised 2 June 2009; accepted 16 August 2009)

This study compares the phoneme recognition performance in speech-shaped noise of a microscopic model for speech recognition with the performance of normal-hearing listeners. "Microscopic" is defined in terms of this model twofold. First, the speech recognition rate is predicted on a phoneme-by-phoneme basis. Second, microscopic modeling means that the signal waveforms to be recognized are processed by mimicking elementary parts of human's auditory processing. The model is based on an approach by Holube and Kollmeier [J. Acoust. Soc. Am. **100**, 1703–1716 (1996)] and consists of a psychoacoustically and physiologically motivated preprocessing and a simple dynamic-time-warp speech recognizer. The model is evaluated while presenting nonsense speech in a closed-set paradigm. Averaged phoneme recognition rates, specific phoneme recognition rates, and phoneme confusions are analyzed. The influence of different perceptual distance measures and of the model's *a-priori* knowledge is investigated. The results show that human performance can be predicted by this model using an optimal detector, i.e., identical speech waveforms for both training of the recognizer and testing. The best model performance is yielded by distance measures which focus mainly on small perceptual distances and neglect outliers.
 [DOI: 10.1121/1.3224721]

PACS number(s): 43.71.An, 43.66.Ba, 43.71.Es, 43.72.Dv [MSS] Pages: 2635–2648

I. INTRODUCTION

The methods usually used for speech intelligibility prediction are index-based approaches, for instance, the articulation index (AI) (ANSI, 1969), the speech transmission index (STI) (Steeneken and Houtgast, 1980), and the speech intelligibility index (SII) (ANSI, 1997). AI and SII use the long-term average frequency spectra of speech and noise separately and calculate an index that can be transformed into an intelligibility score. The parameters used for the calculation are tabulated and mainly fitted to empirical data. These indices have been found to successfully predict speech intelligibility for normal-hearing subjects within various noise conditions and in silence (e.g., Kryter, 1962; Pavlovic, 1987). The STI is also index-based and uses the modulation transfer function to predict the degradation of speech intelligibility by a transmission system. All of these approaches work "macroscopically," which means that macroscopic features of the signal like the long-term frequency *spectrum* or the *signal-to-noise ratios* (SNRs) in different frequency bands are used for the calculation. Detailed *temporal* aspects of speech processing that are assumed to play a major role within our auditory speech perception are neglected. Some recent modifications to the SII improved predictions of the intelligibility in fluctuating noise (Rhebergen and Versfeld, 2005; Rhebergen *et al.*, 2006; Meyer *et al.*, 2007b) and included aspects of temporal processing by calculating the SII based on short-term frequency spectra of speech and noise. However, even these approaches do not mimic all details of auditory preprocessing that are most likely involved in extracting the relevant speech information. Furthermore, the model approaches mentioned above are "macroscopic" in a second sense as they usually predict average recognition rates of whole sets of several words or sentences and not the recognition rates and confusions of single phonemes.

The goal of this study is to evaluate a "microscopic" speech recognition model for normal-hearing listeners. We define microscopic modeling twofold. First, the *particular stages* involved in the speech recognition of normal-hearing human listeners are modeled in a typical way of psychophysics based on a detailed "internal representation" (IR) of the speech signals. Second, the recognition rates and confusions of *single phonemes* are compared to those of human listeners. This definition is in line with Barker and Cooke (2007), for instance. In our study, this kind of modeling is aimed at understanding the factors contributing to the perception of speech in normal-hearing listeners and may be extended to other acoustical signals or to understanding the implications of hearing impairment on speech perception (for an overview see, e.g., Moore, 2003).

Toward this goal we use an auditory preprocessing based on the model of Dau *et al.* (1996a) that processes the signal waveform. This processed signal is then recognized by a dynamic-time-warp (DTW) speech recognizer (Sakoe and Chiba, 1978). This is an approach proposed by Holube and Kollmeier (1996). The novel aspect of this study compared to Holube and Kollmeier (1996) is that the influence of different perceptual distance measures used to distinguish between phonemes within the speech recognizer is investigated in terms of the resulting phoneme recognition scores. Furthermore, we evaluate the predictions of this model on a phoneme scale, which means that we compare confusion ma-

[a)]Parts of this research were presented at the eighth annual conference of the International Speech Communication Association (Interspeech 2007, Antwerp, Belgium).

trices as well as overall speech intelligibility scores. This is a method commonly used in automatic speech recognition (ASR) research.

A. Microscopic modeling of speech recognition

There are different ways to predict speech intelligibility using auditory models. Stadler *et al.* (2007) used an information-theory approach in order to evaluate preprocessed speech information. This approach predicts the speech reception threshold (SRT) very well for subjects with normal hearing for a Swedish sentence test. Another way was presented by Holube and Kollmeier (1996) who used a DTW speech recognizer as a back-end to the auditory model proposed by Dau *et al.* (1996a). They were able to predict speech recognition scores of a rhyme test for listeners with normal hearing and with hearing impairment with an accuracy comparable to that of AI and STI. Both Stadler *et al.* (2007) and Holube and Kollmeier (1996) used auditory models that were originally fitted to other psychoacoustical experiments, such as masking experiments of non-speech stimuli, for instance.

Several studies indicate that temporal information is essential for speech recognition. Chi *et al.* (1999) and Elhilali *et al.* (2003), for instance, compared the predictions of a spectro-temporal modulation index to the predictions of the STI and showed that spectro-temporal modulations are crucial for speech intelligibility. They concluded that information within speech is not separable into a temporal-only and a spectral-only part but that also joint spectro-temporal dimensions contribute to overall performance. Christiansen *et al.* (2006) showed that temporal modulations of speech play a crucial role in consonant identification. For these reasons, this study uses a slightly modified version of the approach by Holube and Kollmeier (1996). The modification is a modulation filter bank (Dau and Kohlrausch, 1997) extending the perception model of Dau *et al.* (1996a), which gives the input for the speech recognition stage. It provides the recognizer with information about the modulations in the different frequency bands. The whole auditory model is based on psychoacoustical and physiological findings and was successful in describing various masking experiments (Dau *et al.*, 1996b), modulation detection (Dau and Kohlrausch, 1997), speech quality prediction (Huber and Kollmeier, 2006), and aspects of timbre perception (Emiroğlu and Kollmeier, 2008). Using a speech recognizer subsequently to the auditory model, as proposed by Holube and Kollmeier (1996), allows for predicting the SRT of an entire speech test. This approach can certainly not account for syntax, semantics, and prosody that human listeners take advantage of. To rule out these factors of human listeners' speech recognition, in the experiments of this study nonsense speech material is presented in a closed response format. The use of this speech material provides a fair comparison between the performance of human listeners and the model (cf. Lippmann, 1997). Furthermore, a detailed analysis of recognition rates and confusions of single phonemes is possible. Confusion matrices can be used in order to compare phoneme recognition rates and phoneme confusions between both humans and model results. Confusion matrices, like those used by Miller and Nicely (1955), can also be used to compare recognition rates between different phonemes provided that systematically composed speech material such as logatoms (short sequences of phonemes, e.g., vowel-consonant-vowel-utterances) is used.

The nonsense speech material of the Oldenburg logatom (OLLO) corpus (Wesker *et al.*, 2005), systematically composed from German vowels and consonants, is used for this task. This corpus was used in a former study (cf. Meyer *et al.*, 2007a) to compare human's speech performance with an automatic speech recognizer. The OLLO speech material in the study of Meyer *et al.* (2007a) allowed excluding the effect of language models that are often used in speech recognizers. Language models store plausible possible words and can use this additional information to crucially enhance the performance of a speech recognizer. Nonsense speech material was also used, for instance, in speech and auditory research to evaluate speech recognition performance of hearing impaired persons (Dubno *et al.*, 1982; Zurek and Delhorne, 1987) and to make a detailed performance comparison between automatic and human speech recognition (HSR) (Sroka and Braida, 2005). Furthermore, nonsense speech material was used, for instance, to evaluate phonetic feature recognition (Turner *et al.*, 1995) and to evaluate consonant and vowel confusions in speech-weighted noise (Phatak and Allen, 2007).

B. *A-priori* knowledge

A model for the prediction of speech intelligibility which uses an internal ASR stage deals with the usual problems of such ASR systems: error rates are much higher than those of normal-hearing human listeners in clean speech (cf. Lippmann, 1997; Meyer and Wesker, 2006) and in noise (Sroka and Braida, 2005; Meyer *et al.*, 2007a). Speech intelligibility models without an ASR stage, e.g., the SII, are provided with more *a-priori* information about the speech signal. The SII "knows" which part of the signal is speech and which part of the signal is noise because it gets them as separate inputs, which is an unrealistic and "unfair" advantage over models using an ASR stage.

For modeling of HSR the problem of too high error rates when using a speech recognizer can be avoided using an "optimal detector" (cf. Dau *et al.*, 1996a) which is also used in many psychoacoustical modeling studies. It is assumed that the recognizing stage of the model after the auditory preprocessing has perfect *a-priori* knowledge of the target signal. Limitations of the model performance are assumed to be completely located in the preprocessing stage. This strategy can be applied to a speech recognizer using template waveforms (for the training of the ASR stage) that are identical to the waveforms of the test signals except for a noise component constraining the performance. Holube and Kollmeier (1996) applied an optimal detector in form of a DTW speech recognizer as a part of their speech intelligibility model using identical speech recordings that were added with different noise passages for the model training stage and for recognizing. Hant and Alwan (2003) and Messing *et al.*

(2008) also used this "frozen speech" approach to model the discrimination of speech-like stimuli. Assuming perfect *a-priori* knowledge using an optimal detector (i.e., using identical recordings as templates and as test items) is one special case of modeling human's speech perception. Another case is using different waveforms for testing and training, thus assuming only limited knowledge about the target signal. This case corresponds not to an optimal detector but to a limited one. The latter is the standard of ASR; the former is widely used in psychoacoustic modeling. In this study, we use both the optimal detector approach and a typical ASR approach. In this way it is possible to investigate how predictions of these two approaches differ and whether the first or the second method is more appropriate for microscopic modeling of speech recognition.

C. Measures for perceptual distances

Because the effects of higher processing stages (like word recognition or use of semantic knowledge) have been excluded in this study by the use of nonsense speech material, it is possible to focus on the sensory part of speech recognition. As a working hypothesis we assume that the central human auditory system optimally utilizes the speech information included in the IR of the speech signal. This information is used to discriminate between the presented speech signal and other possible speech signals. We assume that the auditory system somehow compares the incoming speech information to an internal vocabulary "on a perceptual scale." Therefore, the following questions are of high interest for modeling: what are the mechanisms of comparing speech sounds and what is the best distance measure, on a perceptual scale, for an optimal exploitation of the speech information?

For the perception of musical tones Plomp (1976) compared the perceived similarity of tones to their differences within an equivalent rectangular bandwidth (ERB) sound pressure level spectrogram using different distance measures. Using the absolute value metric, he found higher correlations than using the Euclidean metric. For vowel sounds, however, he found a high correlation using the Euclidean metric. Emiroğlu (2007) also found that the Euclidean distance is more appropriate than, e.g., a cross-correlation measure for comparison of musical tones. The Euclidean distance was also used by Florentine and Buus (1981) to model intensity discrimination and by Ghitza and Sondhi (1997) to derive an optimal perceptual distance between two speech signals. Although the Euclidean distance was preferred by these authors for modeling the perception of sound signals, especially of speech, it still seems to be useful in this study to analyze the differences occurring on the model's "perceptual scale." By using an optimal distance measure, deduced from the empirically found distribution of these differences, the model recognition performance can possibly be optimized.

II. METHOD

A. Model structure

1. The perception model

Figure 1 shows the processing stages of the perception model. The upper part of this sketch represents the training

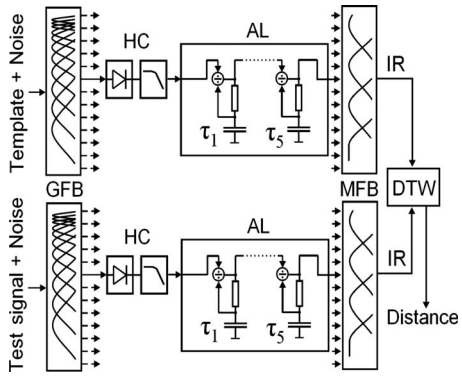


FIG. 1. Scheme of the perception model. The time signals of the template recording added with running noise and the time signal of the test signal added with running noise are preprocessed in the same effective "auditory-like" way. A gammatone filterbank (GFB), a haircell (HC) model, adaptation loops (ALs), and a modulation filterbank (MFB) are used. The outputs of the modulation filterbank are the internal representations (IRs) of the signals. They serve as inputs to the Dynamic-Time-Warp (DTW) speech recognizer that computes the "perceptual" distance between the IRs of the test logatom and the templates.

procedure. A template speech signal with optionally added background noise serves as input to the preprocessing stage. The preprocessing consists of a gammatone-filterbank (Hohmann, 2002) to model the peripheral filtering in the cochlea. 27 gammatone filters are equally spaced on an ERB-scale with one filter per ERB covering a range of center frequencies from 236 Hz to 8 kHz. In contrast to Holube and Kollmeier (1996), gammatone filters with center frequencies from 100 to 236 Hz are omitted because these filters are assumed not to contain information that is necessary to discriminate different phonemes. This is consistent with the frequency channel weighting within the calculation of the SII (ANSI, 1997) and our own preliminary results. A hearing threshold simulating noise that is spectrally shaped to human listeners' audiogram data (according to IEC 60645-1) is added to the signal before it enters the gammatone-filterbank (GFB) (cf. Beutelmann and Brand, 2006). The noise is assumed to be 4 dB above human listeners' hearing threshold for all frequencies, as proposed by Breebaart *et al.* (2001).[1] Each filter output is half-wave rectified and filtered using a first order low pass filter with a cut-off frequency of 1 kHz mimicking a very simple hair cell (HC) model. The output of this HC model is then compressed using five consecutive adaptation loops (ALs) with time constants as given in Dau *et al.* (1996a) (τ_1=5 ms, τ_2=50 ms, τ_3=129 ms, τ_4=253 ms, and τ_5=500 ms). These ALs compress stationary time signals approximately logarithmically and emphasize on- and offsets of non-stationary signals. Furthermore, a modulation filterbank (MFB) according to Dau and Kohlrausch (1997) is used. It contains four modulation channels per frequency channel: one low pass with a cut-off frequency of 2.5 Hz and three band passes with center frequencies of 5, 10, and 16.7 Hz. The bandwidths of the band pass filters are 5 Hz for center frequencies of 5 and 10 Hz, and 8.3 Hz for the band pass with center frequency of 16.7 Hz. The output of this model is an IR that is downsampled to a sampling frequency of 100 Hz. The IR thus contains a two-dimensional feature-matrix at each 10 ms time step consisting of 27 frequency channels and four modulation frequency

channels. The elements of this matrix are given in arbitrary model units (MU). Without the MFB 1 MU corresponds to 1 dB sound pressure level (SPL).

2. The DTW speech recognizer

The IR is passed to a DTW speech recognizer (Sakoe and Chiba, 1978) to "recognize" a speech sample. This DTW can be used either as an optimal detector by using a configuration that contains perfect *a-priori* knowledge or as a limited detector by withholding this knowledge (for details about these configurations see below). The DTW searches for an optimal time-transformation between the IRs of the template and the test signal by locally stretching and compressing the time axes.

The optimal time-transformation between two IRs is computed by first creating a distance matrix D. Each element $D(i,j)$ of this matrix is given by the distance between the feature-matrices of the template's IR ($\mathrm{IR}_{\mathrm{templ}}$) at time index i and the feature-matrix of the test item's IR ($\mathrm{IR}_{\mathrm{test}}$) at time index j. Different distance measures are possible in this procedure (see below). As a next step a continuous "warp path" through this distance matrix is computed (Sakoe and Chiba, 1978). This warp path has the property that averaging the matrix elements along the warp path results in a minimal overall distance. The output of the DTW is this overall distance and thus is a distance between these IRs. From an assortment of possible templates the template with the smallest distance is chosen as the recognized one.

3. Distance measures

In a first approach the Euclidean distance

$$D_{\mathrm{Euclid}}(i,j) = \sqrt{\sum_{f_{\mathrm{mod}}} \sum_{f} \left(\mathrm{IR}_{\mathrm{templ}}(i,f,f_{\mathrm{mod}}) - \mathrm{IR}_{\mathrm{test}}(j,f,f_{\mathrm{mod}})\right)^2} \tag{1}$$

between the feature-vectors $\mathrm{IR}_{\mathrm{templ}}$ and $\mathrm{IR}_{\mathrm{test}}$ was used with f denoting the frequency channel and f_{mod} denoting the modulation-frequency channel of the IRs (Jürgens *et al.*, 2007). In many studies this Euclidean distance is used when comparing perceptual differences (e.g., Plomp, 1976; Holube and Kollmeier, 1996). The Euclidean distance measure implies a Gaussian distribution of the differences between template and test IR.

As an example, Fig. 2 panel 1 shows the normalized histogram of differences Δd between the IRs ($\mathrm{IR}_{\mathrm{templ}}$ and $\mathrm{IR}_{\mathrm{test}}$) of two different recordings of the logatom /ada:/:

$$\Delta d(f,f_{\mathrm{mod}},i,j) = \mathrm{IR}_{\mathrm{templ}}(i,f,f_{\mathrm{mod}}) - \mathrm{IR}_{\mathrm{test}}(j,f,f_{\mathrm{mod}}). \tag{2}$$

In this example, the logatoms were spoken by the same male German speaker and mixed with two passages of uncorrelated ICRA1-noise (Dreschler *et al.*, 2001) at 0 dB SNR. The ICRA1-noise is a steady-state noise with speech-shaped long-term spectrum. Note that Δd corresponds to all differences occurring within a distance matrix, even those that are not part of the final warp path. However, the shape of the histogram is typical of almost all speakers and all SNRs. To investigate the shape of the histogram of differences Δd between these two IRs a Gaussian probability density function (PDF)

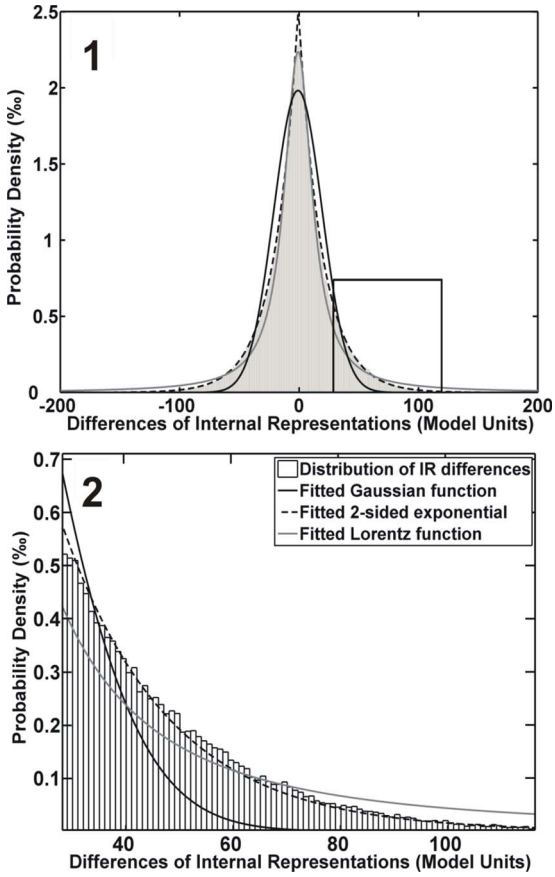


FIG. 2. (Color online) Distribution of differences (in MU) between IRs of two different recordings of the logatom /ada:/. The recordings were spoken by the same male German speaker with "normal" speech articulation style and mixed with ICRA1-noise at 0 dB SNR. A Gaussian, a two-sided-exponential, and a Lorentz-function were fitted to the data, respectively. Panel 1: complete distribution; panel 2: detail (marked rectangular) of panel 1.

$$\mathrm{PDF}_{\mathrm{Gauss}}(\Delta d) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{1}{2}\left(\frac{\Delta d_{\max} - \Delta d}{\sigma}\right)^2\right) \tag{3}$$

is fitted to the distribution which corresponds to the Euclidean metric [Eq. (1)] and a two-sided exponential PDF

$$\mathrm{PDF}_{\mathrm{exp}}(\Delta d) = \frac{1}{2\sigma} \exp\left(\left|\frac{\Delta d_{\max} - \Delta d}{\sigma}\right|\right), \tag{4}$$

and a Lorentzian PDF

$$\mathrm{PDF}_{\mathrm{Lorentz}}(\Delta d) = \frac{1}{\sqrt{2}\pi\sigma} \frac{1}{1 + \frac{1}{2}\left(\frac{\Delta d_{\max} - \Delta d}{\sigma}\right)^2} \tag{5}$$

are also fitted to the distribution, respectively. Two fitting parameters, the width of the fitted curve given by σ and the position of the maximum $\Delta d_{\max}$, must be set. The fits in Fig. 2 panel 1 show that the distribution is almost symmetrical with $\Delta d_{\max}=0$ and that high distances of about 50 MU or more are very much more frequent than expected when assuming Gaussian distributed data. Especially, very high distances of about 80 MU or more (cf. Fig. 2 panel 2) are present in the tail of outliers. The Lorentzian PDF provides a

better fit than the Gaussian function. However, it slightly overestimates the amount of outliers. The two-sided exponential function provides the best fit to the data. The two-sided exponential function is capable of reproducing the shape of the mean peak at 0 MU as well as the shape of the tail of outliers.

By taking the negative logarithm of a PDF [Eqs. (3)–(5)] and summing up the distances across all frequency channels and modulation frequency channels, a distance measure is obtained (cf. Press *et al.*, 1992) that can be used within the speech recognition process. This gives the Euclidean distance metric [Eq. (1)] (for Gaussian distributed data), the absolute value distance metric

$$D_{\mathrm{abs}}(i,j) = \sum_{f_{\mathrm{mod}}} \sum_{f} \left(\left|\mathrm{IR}_{\mathrm{templ}}(i,f,f_{\mathrm{mod}}) - \mathrm{IR}_{\mathrm{test}}(j,f,f_{\mathrm{mod}})\right|\right), \quad (6)$$

and the Lorentzian distance metric

$$D_{\mathrm{Lorentz}}(i,j) = \sum_{f_{\mathrm{mod}}} \sum_{f} \log\left[1 + \tfrac{1}{2}(\mathrm{IR}_{\mathrm{templ}}(i,f,f_{\mathrm{mod}}) - \mathrm{IR}_{\mathrm{test}}(j,f,f_{\mathrm{mod}}))^2\right]. \quad (7)$$

Note that the prefactors that normalize the PDFs are not included within Eqs. (1), (6), and (7) because they represent a constant offset in the distance metric which has no effect on the position of the minimum of the overall distance. The parameter σ is set to 1 MU for simplicity. For Eqs. (1) and (6) the value of σ is not relevant to finding the best warp path through the distance matrix (i.e., solving a constrained minimizing problem). However, in Eq. (7), σ is relevant to finding the best warp path because it cannot be factored out as it can for the Euclidean and the absolute value metric. Choosing σ equal to 1 MU results in a very flat behavior of the distance metric for middle and high distances. Other values of σ in the range from 60 to 0.1 MU showed only minor influence to the performance results in preliminary experiments.

A hypothesis for the present study is that using either Eq. (6) or Eq. (7) instead of the Euclidean distance [Eq. (1)] within the DTW speech recognition process may better account for the characteristic differences of the IRs and may improve matching.

B. Speech corpus

Speech material taken from the OLLO speech corpus (Wesker *et al.*, 2005) is used in this study. The corpus consists of 70 different vowel-consonant-vowel (VCV) and 80 consonant-vowel-consonant (CVC) logatoms composed of German phonemes. The first and the last phoneme of one logatom are the same. The middle phonemes of the logatoms are either vowels or consonants which are listed below (represented with the International Phonetic Alphabet, IPA, 1999).

- Consonants:
  /p/, /t/, /k/, /b/, /d/, /g/, /s/, /f/, /v/, /n/, /m/, /ʃ/, /ts/, /l/
- Vowels:
  /a/, /a:/, /ɛ/, /e/, /ɪ/, /i/, /ɔ/, /o/, /ʊ/, /u/

Consonants are embedded in the vowels /a/, /ɛ/, /ɪ/, /ɔ/, and /ʊ/, respectively, and vowel phonemes are embedded in the consonants /b/, /d/, /f/, /g/, /k/, /p/, /s/, and /t/, respectively.

Most of these logatoms are nonsense in German.[2] The logatoms are spoken by 40 different speakers from four different dialect regions in Germany and by ten speakers from France. The speech material covers several speech variabilities such as speaking rate, speaking effort, different German dialects, accent, and speaking style (statement and question). In the present study, only speech material of one male German speaker with no dialect and with "normal" speech articulation style is used.

C. Test conditions

Calculations with the perception model as well as measurements with human listeners were performed under highly similar conditions.

The same recordings from the logatom corpus were used. The logatoms were arranged into groups in which only the middle phoneme varied. With this group of alternatives a closed testing procedure was performed. This means that both the model and the subject had to choose from identical groups of logatoms. This allowed for a fair comparison of human and modeled speech intelligibility because the humans' semantic and linguistic knowledge had no appreciable influence. Furthermore, it allowed the recognition rates and confusions of phonemes to be analyzed. The speech waveforms were set to 60 dB SPL. Stationary noise with speech-like long-term spectrum (ICRA1-noise, Dreschler *et al.*, 2001) downsampled to a sampling frequency of 16 kHz was added to the recordings and 400 ms prior to the recording. The whole signal was faded in and out using 100 ms Hanning-ramps. After computing the IR of the speech signals as described in Secs. II A and II C, the part of it corresponding to the 400 ms noise prior to the speech signal was deleted. This was done in order to give only the information required for discriminating phonemes to the speech recognizer and not the preceding IR of the preceding background noise.

D. Modeling of *a-priori* knowledge

Two configurations of *a-priori* knowledge of the speech recognizer were realized.

- In configuration A five IRs per logatom calculated from five different waveforms were used as templates. The waveforms were randomly chosen from the recordings of one single male speaker with normal speech articulation style. None of the five waveforms underlying these IRs (the vocabulary) was identical to the tested waveform. The logatom yielding the minimum average distance between the IR of the test sample and all five IRs of the templates was chosen as the recognized one. This limited detector approach mimics a realistic task of automatic speech recognizers because the exact acoustic waveform to be recognized was unknown.
- Model configuration B used a single IR per logatom as template. The waveform of the correct response alternative

was identical to the waveform of the test signal. Thus, the resulting IRs of test signal and the correct response alternative differed only in the added background noise and hearing threshold simulating noise that were uncorrelated in time. In contrast to configuration A, this configuration disregards the natural variability of speech. Thus, it assumes perfect knowledge of the speech template to be matched using the DTW algorithm and corresponds to an optimal detector approach.

The calculation was performed ten times using different passages of background noise and hearing threshold simulating noise according to the individual audiograms of listeners participating in the experiments. The whole calculation took 100 h for configuration A (ten times for 150 logatoms at nine SNR values) and 13 h for configuration B on an up to date standard PC.

E. Subjects

Ten listeners with normal hearing (seven male, three female) aged between 19 and 37 years were employed. Their absolute hearing threshold for pure tones in standard audiometry did not exceed 10 dB hearing level (HL) between 250 Hz and 8 kHz. Only one threshold hearing loss of 20 dB at one audiometric frequency was accepted.

F. Speech tests

The recognition rates of 150 different logatoms were assessed using Sennheiser HDA 200 headphones in a sound-insulated booth. The calibration was performed using a Brüel&Kjaer (B&K) measuring amplifier (Type 2610), a B&K artificial ear (Type 4153), and a B&K microphone (Type 4192). All stimuli were free-field-equalized using an FIR-filter with 801 coefficients and were presented diotically. SNRs of 0, −5, −10, −15, and −20 dB were used for the presentation to human listeners. For each SNR a different presentation order of the 150 logatoms was randomly chosen. For this purpose, the 150 recordings were split into two lists, and the order of presentation of the recordings within the two lists was shuffled. Then all ten resulting lists of all SNRs were randomly interleaved for presentation. Response alternatives for a single logatom had the same preceding and subsequent phoneme (closed test); hence, the subject had to choose either from 10 (CVC) or 14 (VCV) alternatives. The subject was asked to choose the recognized logatom from the list and was asked to guess if nothing was understood. The order of response alternatives shown to the subject was shuffled as well. Before the main measurement all subjects were trained with a list of 50 logatoms.

For characterizing the mean intelligibility scores across all logatoms the model function

$$\Psi(x) = \frac{1-g}{1+\exp(4s(\mathrm{SRT}-L))} + g \tag{8}$$

was fitted to the mean recognition rate (combined for CVCs and VCVs) for each SNR by varying the free parameters SRT and s (slope of the psychometric function at the SRT). The SRT is the SNR at approximately 55% recognition rate

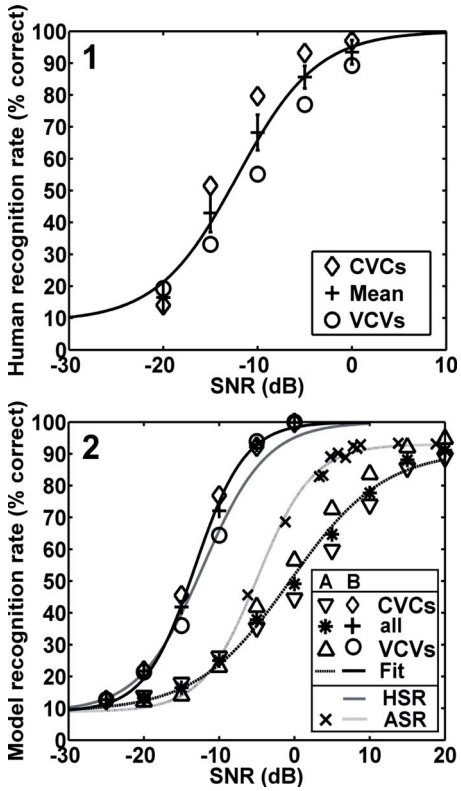


FIG. 3. (Color online) Panel 1: Psychometric function (recognition rate versus SNR) of ten normal-hearing listeners using logatoms in ICRA1-noise. Error bars correspond to the inter-individual standard deviations across subjects. Lines show the fit by Eq. (8). Panel 2: Psychometric function of the perception model with configurations A and B derived with the same utterances of the OLLO speech corpus as for the measurement. The measured psychometric function (taken from panel 1) is additionally shown for comparison as gray line (HSR). For a further comparison, data of Meyer *et al.* (2007a) are plotted (ASR).

(averaged across all CVCs and VCVs) which is the midpoint between the guessing probability and 100%. L corresponds to the given SNR and g is the guessing probability averaged across all CVCs and VCVs ($g=8.9\%$). The fit is performed by maximizing the likelihood assuming that the recognition of each logatom is a Bernoulli trial (cf. Brand and Kollmeier, 2002). Note that this fitting function assumes that 100% recognition rate is reached at high SNRs. This is feasible for listeners with normal hearing and for speech recognition modeling using an optimal detector, but is not necessarily the case for a real ASR system as such an ASR system will still show high error rates on speech material with a low redundancy even when the SNR is very high (Lippmann, 1997). For model configuration A the fitting curve is therefore fixed at the highest recognition rate that occurred in the ASR test.

III. RESULTS AND DISCUSSION

A. Average recognition rates

Figure 3 panel 1 shows the mean phoneme recognition rates in percent correct versus SNR across all phonemes. Error bars denote the inter-individual standard deviations of the ten normal-hearing subjects. Furthermore, the recognition rates of CVCs and VCVs are plotted separately. The recognition rates for CVCs are higher than for VCVs except for −20 dB SNR. The fitting of the psychometric function to

TABLE I. List of fitted parameters characterizing observed and predicted psychometric functions for the discrimination of logatoms in ICRAI noise. Rows denote different distance measures used by the DTW speech recognizer and different model configurations (see Secs. II A and II C for details) as well as values of human listeners. Pearson's rank correlation coefficients (last column) were calculated using the observed data of individual human listeners. * denotes significant ($p<0.05$) and ** highly significant ($p<0.01$) correlations.

	SRT (dB SNR)	Difference to observed SRT (dB)	Slope (%/dB)	Pearson's r^2
Human listeners	−12.2	0[a]	5.4	1[a]
Euclidean, Conf. A	−0.4	11.8	5.7	0.64**
Euclidean, Conf. B	−8.1	4.1	10.0	0.83**
Two-sided exp., Conf. A	−0.4	11.8	5.8	0.65**
Two-sided exp., Conf. B	−10.6	1.6	8.4	0.92**
Lorentzian, Conf. A	−0.6	11.6	3.5	0.83**
Lorentzian, Conf. B	−13.2	−1.0	6.8	0.97**

[a]By definition.

the data yields a slope of $5.4\pm0.6\%$/dB and a SRT of -12.2 ± 1.1 dB. Note that even the recognition rate at −20 dB SNR is significantly above chance and therefore included in the fitting procedure.

The observed and the predicted results calculated with different distance measures and model configurations are shown in Table I. The smallest differences from the observed SRT values are found for configuration B. Using this configuration, the slope of the predicted psychometric function is slightly overestimated. However, model configuration A, which performs a typical task of speech recognizers, shows a large gap of about 12 dB between predicted and observed SRTs, which is typical of ASR (see below). This gap is nearly independent of the type of distance measure, while the slope is slightly underestimated. The last column of Table I shows Pearson's squared rank correlation coefficient r^2 between the individual observed and predicted speech recognition scores. The Lorentzian distance measure using model configuration B shows the highest r^2 of 0.97 ($p<0.01$) whereas the two-sided exponential and the Euclidean distance measure show somewhat lower correlation coefficients and higher differences between observed and predicted SRTs. Different distance measures do not substantially affect the prediction of the SRT using model configuration A.

The predicted psychometric function of this best fitting model realization (configuration B with Lorentzian distance measure) is displayed in Fig. 3 panel 2. In addition, the fitted psychometric function of Fig. 3 panel 1 is replotted (HSR), and the predicted psychometric function of model configuration A with Lorentzian distance measure is shown. Furthermore, ASR-data of Meyer *et al.* (2007a) were included for comparison (see Sec. IV). For model configuration B the resulting SRT using the Lorentzian distance measure is −13.2 dB SNR and thus within the interval of the subjects' inter-individual standard deviation. The ranking of the recognition of vowels and consonants (i.e., that CVCs are better understood than VCVs) is predicted correctly except for −20 dB SNR. Model configuration A, which performs a typical task of speech recognizers, shows a SRT of −0.6 dB and

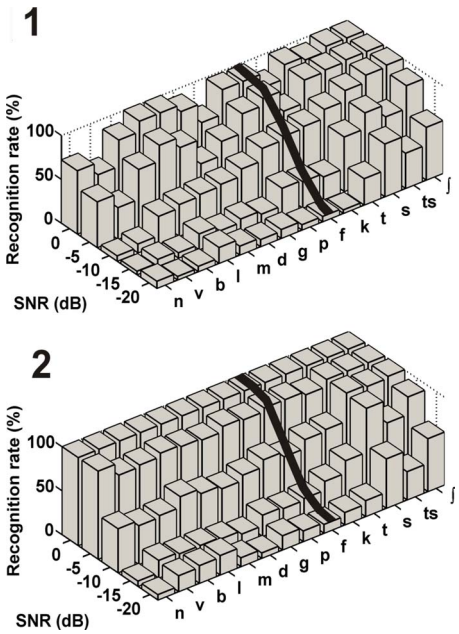


FIG. 4. (Color online) Recognition rates of consonants, separately, as a function of SNR for ten normal-hearing listeners (panel 1) and for model configuration B with Lorentzian distance measure (panel 2). As an example the psychometric function for the discrimination of /f/ in noise is shown (solid line).

a slope of 3.5%/dB using the Lorentzian distance measure. With this configuration the ranking of the recognition of vowels and consonants could not be predicted, i.e., the model shows higher recognition rates for consonants than for vowels.

B. Phoneme recognition rates at different SNRs

Figure 4 shows the recognition rates of single consonants embedded in logatoms as a function of SNR for normal-hearing listeners (panel 1) and for model configuration B using the Lorentzian distance measure (panel 2). Picking out one phoneme, the psychometric function for this specific phoneme can be seen. The solid lines in panels 1 and 2 show these psychometric functions for the phoneme /f/ as an example. Normal-hearing listeners show quite poor recognition rates for the phonemes /n/, /v/, or /g/ at the SNRs chosen for measurement. However, there are also some phonemes like /s/, /ts/, and /ʃ/ that show very high recognition rates at these SNRs. The predicted recognition rates for the latter phonemes (see panel 2) fit the observed recognition rates quite well. This is also the case for /l/, /m/, /p/, /f/, and /t/. For the other phonemes there is a discrepancy between observed and predicted recognition rates especially at high SNRs. For instance, at 0 dB SNR the predicted recognition rate is almost 100% for all phonemes, but normal-hearing listeners actually show poor recognition rates of 58% for /v/ or 70% for /g/. The recognition rates for vowels across SNR are shown in Fig. 5. Normal-hearing listeners show quite a steep psychometric function for the phonemes /e/, /ɛ/, /a:/, and /i/ but a shallower psychometric function for the other phonemes. The predicted recognition rates for /o/ and /u/ fit the observed recognition rates quite well across all SNRs investigated in this study. However, for /e/, /ɛ/, /a:/, and /i/

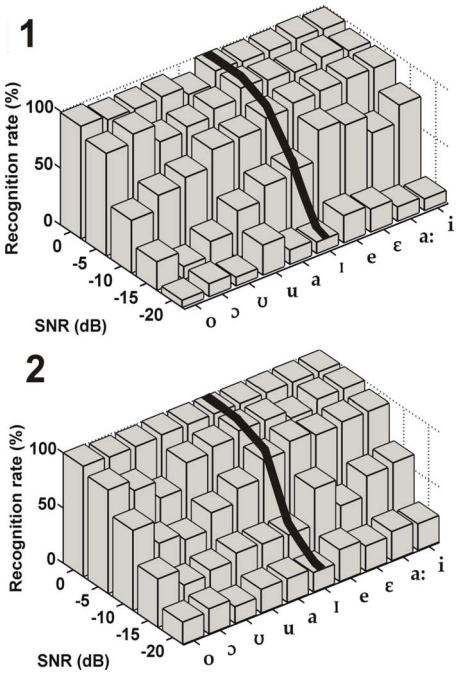


FIG. 5. (Color online) Recognition rates of vowels. The display is the same as in Fig. 4.

the predicted psychometric functions are too shallow. Note that for vowels, contrary to consonants, at 0 dB SNR almost 100% recognition rates are reached by both normal-hearing listeners and model configuration B.

C. Phoneme confusion matrices

Confusion matrices are calculated for all SNRs which were used in the experiment. In Sec. IV the confusion matrices at −15 dB SNR are analyzed. The recognition rates at this SNR are the least influenced by ceiling effects (see Figs. 4 and 5) and show the largest variation across phonemes. Therefore, at this SNR, the patterns of recognition are most characteristic. Figure 6 panel 1 shows the observed confusion matrices of the VCV discrimination task and panel 2 the corresponding predictions using the Lorentzian distance measure with model configuration B. Each row of the confusion matrix corresponds to a specific presented phoneme, and each column corresponds to a recognized phoneme. The diagonal elements denote the rates of correct recognized phonemes and the non-diagonal elements denote confusion rates of phonemes. All numbers are given in percentages.

At −15 dB SNR the average recognition rates for all consonants are 33% (human) and 36% (model configuration B, see also Fig. 3). In the following text the comparison of the two matrices will be described element-wise. Two elements differ significantly if the two-sided 95% confidence intervals surrounding the respective elements do not overlap (cf. Appendix). The observed and the predicted correct consonant recognition rates do not differ significantly, except for the phonemes /s/, /b/, and /v/. Rates below 17% do not differ significantly from the guessing probability of 7% (cf. Appendix). Hence, almost all non-diagonal elements of the model confusion matrix do not differ significantly from the corresponding elements of the human listeners' confusion matrix.

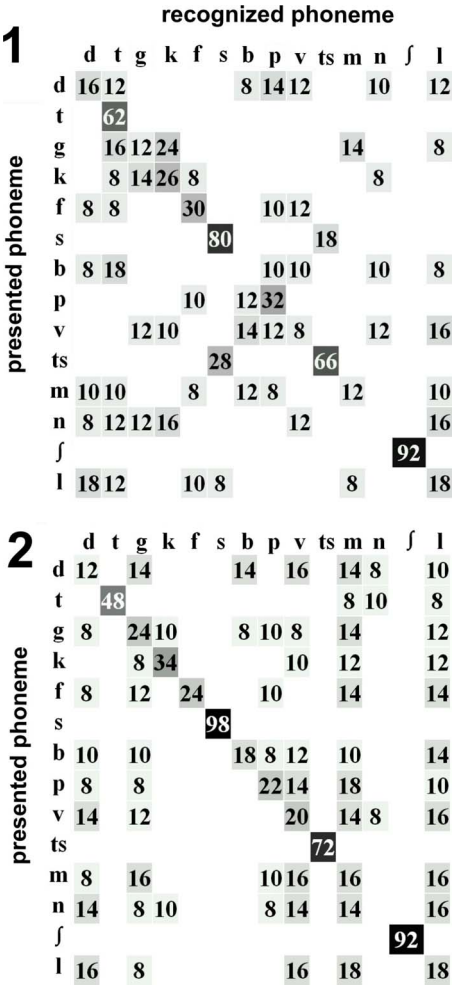


FIG. 6. (Color online) Confusion matrices (response rates in percent) for consonants at −15 dB SNR for normal-hearing subjects (panel 1) and for model configuration B with Lorentzian distance measure (panel 2). Row: presented phoneme; column: recognized phoneme. For better clarity, the values in the cells are highlighted using gray shadings with dark corresponding to high and light corresponding to low response rates. Response rates below 8% are not shown.

One exception is the confusion "presented /ts/-recognized /s/," found in the observed confusion matrix, which cannot be found in the predicted confusion matrix. Other exceptions like "presented /p/-recognized /m/" differ just significantly and shall not be discussed in detail in this section. Unfortunately, the size of confidence intervals of the matrix elements decreases very slowly with an increasing amount of data. Therefore, it is not possible to find many significant differences between predicted and observed matrix elements although the amount of data is already relatively large. However, if we compare the correct recognition rates within one matrix many phonemes can be found that differ significantly in recognition rate. Note that within one single matrix only matrix elements from different rows should be compared (cf. Appendix).

Figure 7 panel 1 shows the observed confusion matrices of the CVC discrimination task and panel B the corresponding predictions using the Lorentzian distance measure with model configuration B. At −15 dB SNR the average recognition rates for all vowels are 52% (human) and 46% (model configuration B, see also Fig. 3 panel 2). The ranking of the

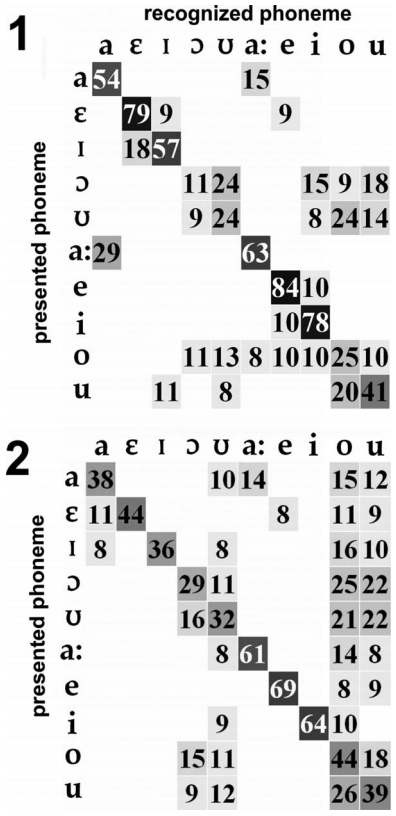


FIG. 7. Confusion matrices (response rates in percent) for vowels at −15 dB SNR for normal-hearing subjects (panel 1) and of model configuration B (panel 2). The display is the same as in Fig. 6.

best recognized phonemes /e/ and /i/, as well as the ranking of the worst recognized phonemes /o/ and /u/, is predicted correctly. However, the overall "contrast" (i.e., the difference between best and worst recognized phonemes) of the predicted matrix is much less pronounced than in the observed matrix. The largest number of confusions occurred between the phonemes /ʊ/, /ɔ/, /o/, and /u/ for both predictions and observations. However, the significant observed confusion "presented /a:/-recognized /a/" cannot be found in the predicted confusion matrix. Furthermore, the phonemes /o/ and /u/ are recognized with a bias by the model, i.e., no matter what phoneme is presented, the model shows a slight preference for these phonemes.

Pearson's φ^2 (Lancaster, 1958) index was used for comparing the similarity between measured and modeled confusion matrix data. This index is based on the chi-square test of equality for two sets of frequencies and provides a normalized measure for the dissimilarity of two sets of frequencies. A value $\varphi^2=1$ is related to complete dissimilarity whereas a value of $\varphi^2=0$ is related to equality. Table II shows φ^2 values for comparing the confusion patterns, i.e., each φ^2 value is a measure for the dissimilarity of the xth row of the observed confusion matrix and the xth row of the predicted confusion matrix of Figs. 6 and 7, respectively. For the consonant confusion matrices highest similarity is found for the confusion patterns of /t/, /s/, and /ʃ/. This very high similarity is mainly due to the high correct response, i.e., the diagonal element. Generally, many observed and predicted confusion patterns show high similarity due to low φ^2-values. However, the observed and predicted confusion patterns of /ts/ show the lowest similarity. This is mainly due to the significant confusion of "presented /ts/-recognized /s/" which was not predicted by the model. The confusion patterns of the phonemes /f/, /l/, and /p/ show moderate similarity. These phonemes also show a poor recognition rate at −15 dB SNR and thus higher percentages in the non-diagonal elements. This gives support to the supposition that the model is not able to predict the consonant *confusions* of normal-hearing listeners. For comparing the patterns of recognition, i.e., the diagonal of the confusion matrix, the correlation coefficients between observed and predicted data are shown in Table III as a function of SNR. For a SNR of −15 dB this correlation coefficient amounts to $r^2=0.91$ ($p<0.01$). This strong correlation means that the model is quite good in modeling the *correct responses*. For observed and predicted consonants there are also highly significant correlations found at −10 and −20 dB SNRs. The correlation decreases rapidly for higher SNR mainly due to ceiling effects, i.e., many phoneme recognition scores are in the range of 100%. Note that at 0 dB SNR a correlation coefficient for consonants could not be assigned due to the fact that at this SNR all consonants are predicted at a recognition rate of 100% whereas some were observed at lower recognition rates.

For the vowel confusion matrices highest similarity is found for the observed and predicted confusion patterns of /a/, /ʊ/, and /u/. Many confusion patterns show a high similarity except for those of /ɛ/, /ɔ/, and /a:/ which show only

TABLE II. Pearson's φ^2 index, a measure of dissimilarity, for comparing the confusion patterns, i.e., one row of a confusion matrix, of observed and predicted phoneme recognitions from Figs. 6 and 7, respectively.

Presented consonant	φ^2	Presented vowel	φ^2
/d/	0.21	/a/	0.10
/t/	0.12	/ɛ/	0.24
/g/	0.24	/ɪ/	0.19
/k/	0.20	/ə/	0.21
/f/	0.16	/ɔ/	0.11
/s/	0.12	/a:/	0.24
/b/	0.15	/e/	0.14
/p/	0.16	/i/	0.15
/v/	0.14	/o/	0.14
/ts/	0.25	/u/	0.10
/m/	0.21		
/n/	0.14		
/ʃ/	0.08		
/l/	0.18		

TABLE III. Correlation coefficients r^2 for comparing observed and predicted recognition scores from Figs. 4 and 5, i.e., the diagonals of confusion matrices, as a function of SNR. * denotes significant ($p<0.05$) and ** highly significant ($p<0.01$) correlations.

SNR (dB)	r^2 for consonants	r^2 for vowels
0	Not assigned	0.09
−5	0.34*	0.52*
−10	0.78**	0.56**
−15	0.91**	0.57*
−20	0.86**	0.26

modest similarity. The high similarity for the former phonemes is mainly due to the correct modeling of confusions "presented /a/-recognized /a:/", "presented /ʊ/-recognized /u/", and "presented /u/-recognized /o/", and the correct responses, respectively. The modest similarity for /ɛ/, /ɔ/, and /a:/ is mainly due to the high discrepancy in predicting the correct diagonal element score. Correlating the diagonals at this SNR (cf. also Table III) shows that the patterns of recognition are significantly ($r^2=0.57$, $p<0.05$) correlated but not as high as for the consonant recognition patterns. This also holds for −10 and −20 dB SNRs. For higher SNRs, i.e., higher average recognition scores, the correlation of predicted and observed vowels is higher than the correlation of consonants. This leads to the assumption that the model can better predict the confusion patterns for vowels than for consonants at low recognition rates as, e.g., for /ʊ/ and /u/. In predicting the correct responses, however, the model is not as good for the vowels as for the consonants.

The fact that the model is not able to predict confusion patterns correctly, especially for consonants, may be due to two reasons. The first reason may be that the model is partly not able to exploit similarities between the IRs of phonemes that might, in fact, be similar to one another for normal-hearing listeners. This is supported by a confusion that is not predicted ("presented /ts/-recognized /s/"), but not, e.g., by the confusions between /u/ and /o/ that are almost correctly predicted. The second reason may be simply due to the high ranges of confidence intervals (see Appendix) due to the inherent binomial statistics of this speech test.

IV. GENERAL DISCUSSION

A. Microscopic prediction of speech intelligibility

This study compares the recognition performance in noise of a microscopic speech intelligibility prediction model to the phoneme recognition performance of human listeners. The model was also used with the same approach as in this study to predict speech intelligibility of a rhyme test (Holube and Kollmeier, 1996). Our results, as well as the results of Holube and Kollmeier (1996), show that this combination of perception model and DTW speech recognizer is able to discriminate noisy speech signals in a closed-set testing procedure. The model used here is also similar to the microscopic model used by Barker and Cooke (2007). Their model is inspired by ASR techniques and evaluates speech parts that "glimpse" the spectro-temporal pattern of the signal to be recognized out of background noise. One main novelty of this study is that the use of the speech database of Wesker *et al.* (2005), which provides many recordings of the same logatom, allows the investigation of the influence of *a-priori* knowledge about the speech. This investigation is possible because the speech recognizer is realized with two model configurations. In model configuration B templates are used which are identical to the test items; this corresponds to maximum *a-priori* knowledge. In model configuration A the recognizer used templates which are not identical to the test items corresponding to less *a-priori* knowledge.

Assuming limited *a-priori* knowledge within model configuration A results in a much poorer performance than observed in the results of human listeners. This reflects the gap between human and machine speech reception (cf. Jürgens *et al.*, 2007) because configuration A is the standard case for ASR. The gap of about 11–12 dB SNR is consistent with findings of other studies employing common speech recognition systems like hidden-Markov-models (HMMs). Meyer *et al.* (2007a) found a gap of about 10 dB SNR (averaged across different speakers) between human listeners' SRT and the SRT of a speech recognizer using mel-frequency-cepstral-coefficients and a HMM using the same OLLO speech corpus and very similar listening experiments. As a direct comparison, a subset of the ASR-data of Meyer *et al.* (2007a) is plotted as an additional psychometric function in Fig. 3. The subset of speech material to be tested is limited to the same speech material that was used in the present study. For this speech material the gap in SRT between ASR and normal-hearing listeners' performance extends to about 8 dB. The difference of 3–4 dB from our results might be due to different speech recognizers used. Meyer *et al.* (2007a) used a speech recognizer that benefited from decades of research. Also the amount of training material in their study was much larger (49 speakers with different articulation styles) than in the present study.

Speech intelligibility can be predicted with greater accuracy using model configuration B in which the amount of information about the speech signal prior to the recognizing process is assumed to be perfect. It has to be stated that in this point the model differs from human listeners' speech processing because human listeners have not stored the exact IR of the signal to be recognized. Human listeners are able to generalize their IR of a speech utterance to different speech waveforms, even if different articulation styles or speakers are involved. However, our speech recognition model includes a pattern recognizer that has to find a speech pattern among different alternatives, which is closer to human speech processing than, for example, the SII (ANSI, 1997). This optimal detector concept is a standard in psychoacoustic modeling and predicts, e.g., forward, backward, and simultaneous masking thresholds (Dau *et al.*, 1996b), modulation detection thresholds (Dau and Kohlrausch, 1997), and the time resolution of the binaural system (Breebaart *et al.*, 2002). As this speech recognition study is in line with other psychoacoustic experiment studies because of the closed-test paradigm and the nonsense speech material used here, such an approach seems to be appropriate. The very accurate agreement of observed and predicted phoneme recognition rates using model configuration B does not mean that human listeners have a perfect decision device. Humans' limitations in discriminating speech in noise are certainly due to energetic masking of the speech signal by background noises and also due to errors in the inherent processing in the subsequent word recognition stage. However, the speech discrimination performance of the model is very similar to that of human listeners if all limitations of performance are assumed entirely in the preprocessing stage of the model. For the experiments presented here this may be interpreted as that lifelong training of humans in speech makes the pattern recognizing part of HSR perform as well as the model's optimal detector.

With configuration B the model is capable of predicting the SRT of this speech test with an accuracy of about 1 dB. The SII (ANSI, 1997) predicts the SRT within the same accuracy range: For −15 dB SNR the SII-value is found to be 0.045, for instance, and for −10 dB the SII is 0.18. Transformed to intelligibility scores by using the SII transfer function for Hagerman's sentences in noise (Magnusson, 1996), the resulting SRT is −11.2 dB SNR. The main advantage of the microscopic modeling approach compared to the SII is that, whereas the SII is able to predict only average recognition scores, this approach is able to predict the recognition scores for each phoneme separately. Furthermore, this approach draws out some characteristic phoneme confusions that are commonly seen.

B. Distance measures

The type of distance measure crucially influences the performance of the speech recognizer when using model configuration B. The Euclidean distance used by, e.g., Plomp (1976), Holube and Kollmeier (1996), and Jürgens *et al.* (2007) shows the poorest performance among the distance measures investigated here. In this study, there is a gap of more than 4 dB between the SRT of model configuration B and human listeners' SRT. Using the Euclidean distance, outlying passages are strongly weighted and consequently the DTW algorithm tries to minimize the occurrence of outlying passages as far as possible. This may cause the warp path, i.e., the temporal matching function between two IRs, to be fitted more to the passages containing different speech or noise. Passages with low distances are disregarded. By applying a distance measure that is less sensitive to outliers in the matching procedure of two IRs (i.e., using the two-sided exponential measure or the Lorentzian measure) this gap is substantially decreased or vanishes. Using the two-sided exponential distance measure, all distances are weighted with their usual occurrence probability (cf. Fig. 2). Therefore, this can be called a "natural" distance measure for speech in noise. Although no substantial influence of the type of distance measure was found on the performance of model configuration A, it was found for model configuration B. One could argue, since configuration A is typical of an ASR system, that other ASR systems may not benefit from an optimization of the distance measure they use. However, as this approach uses a speech recognizer that does not require a large amount of training material as common ASR systems do, this is speculative. Nevertheless, for further optimizing of ASR systems it may be useful to study the influence of different distance measures on the ASR systems' performance.

Using the Lorentzian distance measure, all outlying passages get approximately the same constant weight because of the flatness of the logarithm for large input values. Therefore, the overall distance between two IRs is mainly dominated by the smallest elements of the distance matrix. In other words, the steepness of the logarithm at low values causes similar passages of the IRs to be matched as closely as possible. This may particularly be an advantage for discriminating noisy speech samples because the speech recognizer is dominated by matched (i.e., similar speech) patterns and neglects unmatched (i.e., noise or different speech) patterns. Hence, the detector can separate the objects "matched speech" passages from "unmatched speech" or "noise only" passages more appropriately. If we conceive of noise and speech as different acoustical objects this mechanism may have some similarities to the mechanism of acoustical object separation within the human auditory system. Neglecting passages that do not match passages of stored response alternatives is a candidate for modeling human's mechanism of object separation. In that way the distinction between a "matchable speech object" and a "not matchable speech object" or "noise-only object" may be enhanced. Using model configuration B, the Lorentzian distance measure performs best and results in a high agreement in phoneme recognition. Therefore, this set-up was chosen for the prediction of speech recognition in noise of listeners with normal hearing.

C. Phoneme recognition rates and confusions

In this study both human listeners and the model show the highest performance at the same consonants /t/, /s/, /ʃ/, and /ts/ as in the study of Phatak and Allen (2007) who investigated consonant recognition rates in speech weighted noise. The results obtained in this study are in line with those of Phatak and Allen (2007), although they used speakers and listeners of a different native language and different speech material. Furthermore, the amount of alternatives that could be recognized was completely different from our measurements. However, the separation of consonants into a low scoring and a midscoring group with the same phonemes as in Phatak and Allen (2007) could not be observed in this study. They concluded that differences in recognition rates can mainly be explained by differences in the long-term spectra of speech and noise. However, this may not account for consonants with characteristics that are mainly determined by the temporal structure as, e.g., for plosives like /p/, /t/, or /k/. Our approach regards this temporal structure by the temporal matching performed in the DTW speech recognizer.

By and large, the confusion matrices of human listeners and of model configuration B with Lorentzian distance measure are very similar. Except for a small number of elements, the consonant confusion matrices do not differ significantly element-wise regarding the binomial statistics valid for these discrimination tasks (see Appendix). The correlation between predicted and observed recognition rates of single phonemes is very high. This is promising and it may indicate that for all phonemes speech information is conserved or emphasized during the modeled "effective" auditory preprocessing in a way similar to human listeners.

The vowel confusion matrix of the model shows a slight preference, i.e., a bias, concerning the vowels /ʊ/, /ɔ/, /o/, and /u/ independent of the presented vowel. This is one main difference between the predicted and observed vowel confusion matrices. Meyer *et al.* (2007a) found that the phonemes /o/ and /u/ within this speech corpus have the least distinctive average spectrum compared to speech-shaped noise. Consequently these phonemes are the phonemes best masked in the background noise at low SNRs. If the speech recognizer is

not able to match a presented phoneme, it is very probable that it matches the IR that is the most similar to the IR of the background noise. These are the IR of logatoms with /o/ and /u/ as middle phonemes. In some cases the procedure probably matches mainly the background noise characteristics of the IR and is not able to focus on the speech characteristics anymore. One reason why the prediction of vowel recognition rates is poorer than for consonants while the prediction of vowel confusions is better than for the consonants may be the spectro-temporal structure of these two phoneme groups. Generally, vowels are more stationary signals than consonants. Furthermore, there is no clear separation between different vowels but a continuous transition in the frequency range. Therefore, it seems reasonable to assume that two different vowels are "perceptually" closer to one another than are two different consonants. This may explain why confusions occur more frequently in both normal-hearing listeners' and modeled data.

D. Variability in the data

Data obtained by speech tests using human listeners always show both intra-individual and inter-individual variabilities. One factor for the inter-individual variability is the variability of the hearing threshold across listeners. Preliminary simulations, however, showed that adapting only the hearing threshold simulating noise results in less variability than found in normal-hearing listeners' speech recognition data. This can be explained by the low rms level of the hearing threshold simulating noise which is masked by the much higher level of the background noise. For this reason a much more effective way to include variability was to use running background noise. In other words the variability in the simulations originates almost exclusively from the statistics of the background noise. However, this is somewhat unrealistic because in the measurements the background noise stimuli were identical for every participant whereas, in reality the auditory processing varied. It still remains an open question how to obtain a comparable variability by modifying the auditory processing without using this workaround. For speech intelligibility modeling in silence, e.g., Holube and Kollmeier (1996) achieved some variability using a fluctuating absolute threshold of hearing which improved their predictions in silence. Due to the small influence of the exact form of the absolute hearing threshold in our study, this procedure was not applied here.

E. Practical relevance

There are at least two different applications that may benefit from this modeling approach. First, this approach may be used to model sensorineural hearing loss by appropriate manipulation of the auditory preprocessing. Hence, consequences of the auditory preprocessing on speech recognition for listeners with impaired hearing can be investigated. As a long-term aim the model may serve as a tool for distinguishing between reduced speech recognition caused by impaired preprocessing or by further problems in the patient's central processing. A further long-term aim is to find processing strategies that substantially enhance the recognition performance of certain phonemes and that can be used in hearing-aids. Second, automatic speech recognizers may be improved especially for functioning in noise, if they focus on passages fitting well to their vocabulary and if they neglect outlying passages in a manner similar to that used in the weighting of the perceptual distance in this study.

V. CONCLUSIONS

(1) The microscopic approach for predicting speech intelligibility by using an auditory model as a pre-processor to a DTW speech recognizer is capable of discriminating CVC and VCV logatoms in noise.

(2) If the detector stage is assumed to be optimal by using identical templates for test signal and vocabulary, the speech discrimination performance of the model is very similar to that of human listeners. This means that the recognition of logatoms by humans can be modeled effectively by assuming a limited auditory-like preprocessing stage and a perfect speech matching process. In other words, the prediction of normal-hearing listeners' speech recognition is only possible if exactly the same stimulus is available as *a-priori* knowledge.

(3) No substantial improvement in performance of the model with *imperfect* knowledge about the speech signal was found when changing the distance measure.

(4) For the model with *perfect* knowledge about the speech signal, the Lorentzian measure is the best distance measure where outlying passages have the smallest weight compared to the other distance measures such as the Euclidean or the two-sided-exponential.

(5) Predicted recognition rates of each single phoneme are very similar to observed recognition rates but some of the observed characteristic patterns of human confusions did not occur within the predictions.

ACKNOWLEDGMENTS

We thank Birger Kollmeier for his substantial support and contribution to this work and Bernd Meyer for making available the ASR data. Thanks to Mitchell Sommers, Amy Beeston, and one anonymous reviewer who helped to greatly improve the manuscript. We would also like to thank the EU HearCom Project, the "Förderung wissenschaftlichen Nachwuchses des Landes Niedersachsen" (FwN), and SFB/TR 31 "Das aktive Gehör" (URL: http://www.uni-oldenburg.de/sfbtr31) for funding the research reported in this paper.

APPENDIX: SIGNIFICANCE OF CONFUSION MATRIX ELEMENTS

For deciding whether or not two matrix elements differ significantly, a statistical analysis has to be made. One element of a confusion matrix is given by $p=x/n$, with x denoting the number of recognitions of the phoneme specified by the column and n denoting the number of presentations specified by the row of the matrix. There are $n=50$ (VCV) and $n=80$ (CVC) presentations, respectively, of each phoneme at each SNR (i.e., each confusion matrix). Each single presentation is followed by a subjects' decision for one response alternative given in the list. Therefore, each decision

is a Bernoulli-trial with an unknown underlying probability π for the correct item and $(1-\pi)$ for all other items. Note that p is just an estimate of π. By estimating π using p, both-sided 95%-confidence intervals can be calculated based on binomial statistics (Sachs, 1999). The upper boundary is given by

$$\pi_{\text{upper}} = \frac{(x+1)F_{\text{upper}}}{n-x+(x+1)F_{\text{upper}}}, \tag{A1}$$

with $F_{\text{upper}}=F_{\{2(x+1),2(n-x)\}}$ taken from Fisher's F-distribution. The lower boundary is given by

$$\pi_{\text{lower}} = \frac{x}{x+(n-x+1)F_{\text{lower}}}, \tag{A2}$$

with $F_{\text{lower}}=F_{\{2(n-x+1),2x\}}$.

The range of confidence intervals for an observed percentage p in the speech test, i.e. $(\pi_{\text{upper}}-\pi_{\text{lower}})$, results in 4% to 22% for $n=80$ (CVC presentation) and 6% to 29% for $n=50$ (VCV presentation) whereas the wider range can be found at $p=50\%$ and the smaller range at $p=0\%$ and $p=100\%$. These confidence intervals contain the underlying probability π with a confidence of 95%. Furthermore, they offer a criterion to decide if two percentages that are statistically independent of each other differ significantly (i.e., their confidence intervals must not overlap). The precondition, statistical independence within one confusion matrix, is warranted only for two matrix elements that are not part of the same row because in this case completely different phonemes were presented to obtain the two percentages. Two elements of the same row are not independent of each other because the recognition of one phoneme affects the percentages for the other phonemes of that row. A comparison of two elements being part of the same row requires a different statistical analysis that is not discussed here. Therefore, only elements of different rows (or different confusion matrices) can be tested for difference using the methods described in this section. When comparing two different confusion matrices (e.g., observed with predicted) this problem does not occur.

[1]Breebaart *et al.* (2001) found out that a 9.4 dB SPL Gaussian noise within one gammatone filter channel just masks a sinusoid with 2 kHz frequency at absolute hearing threshold (5 dB SPL, which is about 4 dB lower). This approach was extrapolated for other audiometric frequencies.

[2]Even if very few logatoms in this corpus are forenames or may have a meaning in certain dialect regions in Germany these logatoms are not excluded in this study to preserve the very systematic composition of this speech corpus.

ANSI (**1969**). "ANSI S3.5-1969 American national standard methods for the calculation of the articulation index," Standards Secretariat, Acoustical Society of America.

ANSI (**1997**). "ANSI S3.5-1997 Methods for calculation of the speech intelligibility index," Standards Secretariat, Acoustical Society of America.

Barker, J., and Cooke, M. (**2007**). "Modelling speaker intelligibility in noise," Speech Commun. **49**, 402–417.

Beutelmann, R., and Brand, T. (**2006**). "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 331–342.

Brand, T., and Kollmeier, B. (**2002**). "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," J. Acoust. Soc. Am. **111**, 2801–2810.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**). "Binaural processing model based on contralateral inhibition. I. Model structure," J. Acoust. Soc. Am. **110**, 1074–1088.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2002**). "A time-domain binaural signal detection model and its predictions for temporal resolution data," Acta. Acust. Acust. **88**, 110–112.

Chi, T. S., Gao, Y. J., Guyton, M. C., Ru, P. W., and Shamma, S. (**1999**). "Spectro-temporal modulation transfer functions and speech intelligibility," J. Acoust. Soc. Am. **106**, 2719–2732.

Christiansen, T. U., Dau, T., and Greenberg, S. (**2006**). "Spectro-temporal processing of speech—An information-theoretic framework," in *International Symposium on Hearing 2006*, Cloppenburg, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. M. Verhey (Springer, Heidelberg), pp. 517–524.

Dau, T., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulation I. Detection and masking with narrowband-carriers," J. Acoust. Soc. Am. **102**, 2893–2905.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996a**). "A quantitative model of the "effective" signal processing in the auditory system: I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996b**). "A quantitative model of the "effective" signal processing in the auditory system: II. Simulations and measurements," J. Acoust. Soc. Am. **99**, 3623–3631.

Dreschler, W. A., Verschuure, H., Ludvigsen, C., and Westermann, S. (**2001**). "ICRA noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment," Audiology **40**, 148–157.

Dubno, J. R., Dirks, D. D., and Langhofer, L. R. (**1982**). "Evaluation of hearing-impaired listeners using a nonsense-syllable test. 2. Syllable recognition and consonant confusion patterns," J. Speech Hear. Res. **25**, 141–148.

Elhilali, M., Chi, T., and Shamma, S. A. (**2003**). "A spectro-temporal modulation index (STMI) for assessment of speech intelligibility," Speech Commun. **41**, 331–348.

Emiroğlu, S. (**2007**). "Timbre perception and object separation with normal and impaired hearing," Ph.D. thesis, Universität Oldenburg, Oldenburg, Germany.

Emiroğlu, S., and Kollmeier, B. (**2008**). "Timbre discrimination in normal-hearing and hearing-impaired listeners under different noise conditions," Brain Res. **1220**, 199–207.

Florentine, M., and Buus, S. (**1981**). "An excitation-pattern model for intensity discrimination," J. Acoust. Soc. Am. **70**, 1646–1654.

Ghitza, O., and Sondhi, M. M. (**1997**). "On the perceptual distance between speech segments," J. Acoust. Soc. Am. **101**, 522–529.

Hant, J. J., and Alwan, A. (**2003**). "A psychoacoustic-masking model to predict the perception of speech-like stimuli in noise," Speech Commun. **40**, 291–313.

Hohmann, V. (**2002**). "Frequency analysis and synthesis using a gammatone filterbank," Acta. Acust. Acust. **88**, 433–442.

Holube, I., and Kollmeier, B. (**1996**). "Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model," J. Acoust. Soc. Am. **100**, 1703–1716.

Huber, R., and Kollmeier, B. (**2006**). "PEMO-Q—A new method for objective: Audio quality assessment using a model of auditory perception," IEEE Trans. Audio, Speech, Lang. Process. **14**, 1902–1911.

IPA (**1999**). *The Handbook of the International Phonetic Association* (Cambridge University Press, Cambridge), pp. 194–197.

Jürgens, T., Brand, T., and Kollmeier, B. (**2007**). "Modelling the human-machine gap in speech reception: Microscopic speech intelligibility prediction for normal-hearing subjects with an auditory model," in Interspeech 2007, Antwerp, Belgium, pp. 410–413.

Kryter, K. D. (**1962**). "Methods for calculation and use of articulation index," J. Acoust. Soc. Am. **34**, 1689–1697.

Lancaster, H. O. (**1958**). "The structure of bivariate distributions," Ann. Math. Stat. **29**, 719–736.

Lippmann, R. P. (**1997**). "Speech recognition by machines and humans," Speech Commun. **22**, 1–15.

Magnusson, L. (**1996**). "Speech intelligibility index transfer functions and speech spectra for two Swedish speech recognition tests," Scand. Audiol. **25**, 59–67.

Messing, D., Delhorne, L., Bruckert, E., Braida, L., and Ghitza, O. (**2008**). "Consonant discrimination of degraded speech using an efferent-inspired closed-loop cochlear model," in Interspeech 2008, Brisbane, Australia, pp. 1052–1055.

Meyer, B., and Wesker, T. (**2006**). "A human-machine comparison in speech recognition based on a logatom corpus," in Workshop on Speech Recognition and Intrinsic Variation, Toulouse, France.
Meyer, B., Brand, T., and Kollmeier, B. (**2007a**). "Phoneme confusions in human and automatic speech recognition," in Interspeech 2007, Antwerp, Belgium, pp. 1485–1488.
Meyer, R. M., Kollmeier, B., and Brand, T. (**2007b**). "Predicting speech intelligibility in fluctuating noise," in Eighth EFAS Congress Joint Meeting with the Tenth Congress of the German Society of Audiology, Heidelberg, Germany.
Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.
Moore, B. C. J. (**2003**). "Speech processing for the hearing-impaired: Successes, failures and implications for speech mechanisms," Speech Commun. **41**, 81–91.
Pavlovic, C. V. (**1987**). "Derivation of primary parameters and procedures for use in speech intelligibility predictions," J. Acoust. Soc. Am. **82**, 413–422.
Phatak, S. A., and Allen, J. B. (**2007**). "Consonant and vowel confusions in speech-weighted noise," J. Acoust. Soc. Am. **121**, 2312–2326.
Plomp, R. (**1976**). *Aspects of Tone Sensation* (Academic, London).
Press, W., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (**1992**). *Numerical Recipes in C* (Cambridge University Press, Cambridge).
Rhebergen, K. S., and Versfeld, N. J. (**2005**). "A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," J. Acoust. Soc. Am. **117**, 2181–2192.
Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2006**). "Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise," J. Acoust. Soc. Am. **120**, 3988–3997.
Sachs, L. (**1999**). *Angewandte Statistik* (Springer, Berlin).
Sakoe, H., and Chiba, S. (**1978**). "Dynamic programming algorithm optimization for spoken word recognition," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-26**, 43–49.
Sroka, J. J., and Braida, L. D. (**2005**). "Human and machine consonant recognition," Speech Commun. **45**, 401–423.
Stadler, S., Leijon, A., and Hagerman, B. (**2007**). "An information theoretic approach to predict speech intelligibility for listeners with normal and impaired hearing," in Interspeech 2007, Antwerp, Belgium, pp. 389–401.
Steeneken, H. J. M., and Houtgast, T. (**1980**). "Physical method for measuring speech-transmission quality," J. Acoust. Soc. Am. **67**, 318–326.
Turner, C. W., Souza, P. E., and Forget, L. N. (**1995**). "Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners," J. Acoust. Soc. Am. **97**, 2568–2576.
Wesker, T., Meyer, B., Wagener, K., Anemüller, J., Mertins, A., and Kollmeier, B. (**2005**). "Oldenburg logatom speech corpus (OLLO) for speech recognition experiments with humans and machines," in Interspeech 2005, Lisboa, Portugal, pp. 1273–1276, freely available at http://sirius.physik.uni-oldenburg.de (Last viewed 9/11/2009).
Zurek, P. M., and Delhorne, L. A. (**1987**). "Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment," J. Acoust. Soc. Am. **82**, 1548–1559.

Perceptual learning of time-compressed and natural fast speech

Patti Adank[a)]
School of Psychological Sciences, University of Manchester, Manchester, M13 9PL, United Kingdom and Donders Institute for Brain, Cognition and Behaviour, Centre for Cognitive Neuroimaging, Radboud University Nijmegen, 6525 EN, Nijmegen, The Netherlands

Esther Janse
Utrecht Institute of Linguistics, OTS, Utrecht University, 3512 BL, Utrecht, The Netherlands and Max Planck Institute for Psycholinguistics, 6500 AH, Nijmegen, The Netherlands

(Received 18 February 2009; revised 29 June 2009; accepted 6 August 2009)

Speakers vary their speech rate considerably during a conversation, and listeners are able to quickly adapt to these variations in speech rate. Adaptation to fast speech rates is usually measured using artificially time-compressed speech. This study examined adaptation to two types of fast speech: artificially time-compressed speech and natural fast speech. Listeners performed a speeded sentence verification task on three series of sentences: normal-speed sentences, time-compressed sentences, and natural fast sentences. Listeners were divided into two groups to evaluate the possibility of transfer of learning between the time-compressed and natural fast conditions. The first group verified the natural fast before the time-compressed sentences, while the second verified the time-compressed before the natural fast sentences. The results showed transfer of learning when the time-compressed sentences preceded the natural fast sentences, but not when natural fast sentences preceded the time-compressed sentences. The results are discussed in the framework of theories on perceptual learning. Second, listeners show adaptation to the natural fast sentences, but performance for this type of fast speech does not improve to the level of time-compressed sentences.
 [DOI: 10.1121/1.3216914]

PACS number(s): 43.71.Es, 43.71.Bp, 43.71.Gv [PEI] Pages: 2649–2659

I. INTRODUCTION

Within a given conversation, speakers often vary their speech rate considerably (Miller *et al.*, 1984b), ranging between 140 and 180 words/min. These on-line changes in speaking rate affect qualitative aspects of speech: at higher rates, speech is produced with generally more coarticulation and assimilation (Browman and Goldstein, 1990; Byrd and Tan, 1996) sometimes even leading to deletion of segments (Ernestus *et al.*, 2002; Koreman, 2006). Moreover, people increase their speech rate in a nonlinear fashion: higher speaking rates generally affect consonant durations less than vowel durations (Lehiste, 1970; Max and Caruso, 1997). In addition, durations of unstressed syllables in polysyllabic words are reduced more than stressed syllables (Peterson and Lehiste, 1960). These phonetic and phonological consequences of the variations in speaking rate pose a potential problem for listeners, forcing them to constantly normalize for varying speech rate (Green *et al.*, 1994; Miller *et al.*, 1984a; Miller and Liberman, 1979).

Apart from these latter studies on local rate effects on phonetic perception of specific phoneme contrasts, there is a body of research on more gradual adaptation to artificially time-compressed speech. Artificial time compression is a method for artificially shortening the duration of an audio signal without affecting the fundamental frequency of the signal (Golomb *et al.*, 2007; Pallier *et al.*, 1998; Sebastián-Gallés *et al.*, 2000; Wingfield *et al.*, 2003). Listeners can adapt to sentences compressed up to 38% of their original duration within 10–20 sentences (Dupoux and Green, 1997). Adaptation to this manipulation is not immediate, but takes place during exposure to a number of sentences that are initially of very poor intelligibility. While adaptation to time-compressed speech has provided useful insights on general adaptation processes in speech comprehension, it is questionable whether time-compressed speech itself provides a useful model for adaptation to the specific characteristics of naturally produced fast speech. First of all, there is evidence that natural fast speech is more difficult to process than speech that is artificially time compressed to the same rate (Janse, 2004). Second, modern time-compression algorithms (Moulines and Charpentier, 1990) do not significantly affect the long-term spectral characteristics of the original speech signal, while allowing for careful manipulation of the temporal characteristics. Natural fast speech, on the other hand, differs from speech delivered at a normal speaking rate in both spectral and temporal characteristics (Koreman, 2006; Wouters and Macon, 2002).

The aims of the present study were twofold. First, we wanted to establish if listeners adapt to naturally produced fast speech and if so, how this adaptation process compares to adaptation to time-compressed speech. While adaptation to time-compressed speech is usually determined with participants reporting keywords in a sentence (Dupoux and Green, 1997; Golomb *et al.*, 2007; Pallier *et al.*, 1998), adaptation in the present study was measured using reaction times and percent correct as (i) they were expected to pro-

[a)]Author to whom correspondence should be addressed. Electronic mail: patti.adank@manchester.ac.uk

TABLE I. Mean percent correct plus standard deviations (stddev) for both groups for the block 1 three speech types for the ten blocks of six sentences.

		Normal		Time compressed		Natural fast	
% correct		Mean	Stddev	Mean	Stddev	Mean	Stddev
Group 1	Block 1	94.2	23.5	96.7	18.0	71.7	45.3
	Block 2	97.5	15.7	95.0	21.9	80.0	40.2
	Block 3	95.0	21.9	94.2	23.5	75.8	43.0
	Block 4	97.5	15.7	98.3	12.9	75.8	43.0
	Block 5	95.0	21.9	92.5	26.4	69.2	46.4
	Block 6	98.3	12.9	95.0	21.9	81.7	38.9
	Block 7	98.3	12.9	97.5	15.7	82.5	38.2
	Block 8	99.2	9.1	92.5	26.4	85.0	35.9
	Block 9	98.3	12.9	93.3	25.0	84.2	36.7
	Block 10	98.3	12.9	95.8	20.1	88.3	32.2
Group 2	Block 1	97.6	15.3	90.5	29.5	92.1	27.1
	Block 2	98.4	12.5	97.6	15.3	92.1	27.1
	Block 3	100.0	0.0	98.4	12.5	83.3	37.4
	Block 4	98.4	12.5	92.1	27.1	74.6	43.7
	Block 5	88.9	31.6	96.0	19.6	78.6	41.2
	Block 6	98.4	12.5	99.2	8.9	77.8	41.7
	Block 7	98.4	12.5	99.2	8.9	91.3	28.3
	Block 8	97.6	15.3	93.7	24.5	86.5	34.3
	Block 9	97.6	15.3	96.0	19.6	89.7	30.5
	Block 10	99.2	8.9	97.6	15.3	86.5	34.3

vide a more fine-grained measure than percent correctly reported keywords only and (ii) to avoid ceiling effects in the time-compressed speech condition. Speech can be highly time compressed before identification scores of listeners drop below ceiling level. Such fast rates of speech can hardly be attained by humans speeding up their speech rate, which makes it difficult to compare the two types of fast speech at the same rate. Also, Clarke and Garrett (2004) used reaction times to show adaptation to foreign-accented speech. We therefore used a speeded sentence verification task to monitor the adaptation process. This task is based on the Speed and Capacity of Language Processing Test, or SCOLP (Baddeley *et al.*, 1992) of which an aural version was previously described in May *et al.*, 2001; Adank *et al.*, 2009. SCOLP is originally a written test in which the participant verifies as many sentences as possible in 2 min. The sentences are all obviously true or false and all consist of a mismatch of subject and predicate from true sentences (e.g., *Tomato soup is a liquid* versus *Tomato soup is people*). Overall, it provides a sensitive and reliable measure of the speed of language comprehension. When transformed to a speeded verification task, it can be used to determine the cognitive processing cost of a specific task or process, as demonstrated by Adank *et al.* (2009), who used the task to determine the relative cognitive load of comprehension of regionally accented sentences versus sentences in the standard language in noise. A decrease in the speed of processing after exposure to time-compressed speech can thus be taken to signal perceptual learning of the acoustic consequences of, for instance, time compressed or naturally fast speech. We created a Dutch version of the SCOLP sentences as the experiment was run in The Netherlands, with Dutch listeners. Like the English version, the Dutch version was made up of sentences that consisted of a noun plus predicate. A total of 90 sentence pairs were constructed (90 true and 90 false). For example, "Tomaten groeien aan planten" (*tomatoes grow on plants*) as a true sentence and "Tomaten hebben sterke tanden" (*tomatoes have strong teeth*) as a false sentence. All sentences of the Dutch version designed for the present study are listed in the Appendix.

Second, we aimed to establish whether there is transfer of learning in the adaptation process between naturally fast and time-compressed speech: does exposure to time-compressed speech before being exposed to naturally fast speech affect the adaptation process (and vice versa)? Transfer of learning involves the application of skills or knowledge learned in one context to another context (Cormier and Hagman, 1987; Haskell, 2001; Thorndike and Woodforth, 1901). Transfer of learning has been found in the auditory domain for nonspeech stimuli (Delhommeau *et al.*, 2005; Delhommeau *et al.*, 2002) and speech stimuli (Bradlow and Bent, 2008; McClaskey *et al.*, 1983; Tremblay *et al.*, 1997). Transfer of learning was, for instance, reported for auditory frequency discrimination tasks: Delhommeau *et al.* (2002) measured listeners' frequency discrimination thresholds (FDTs) (the smallest audible difference frequency, Δf, around a center frequency) for four center frequencies (750, 1500 3000, and 6000 Hz) before and after training. Listeners were then trained for a specific center frequency (e.g., 750 Hz) and then subsequently tested again at all four center frequencies. Delhommeau *et al.* (2002) found that training at a specific frequency lowered FDTs for that frequency and that the improvement transferred to the other (untrained) frequencies. Furthermore, McClasky *et al.* (1983) trained listeners to perceive prevoiced labial syllables and found that they

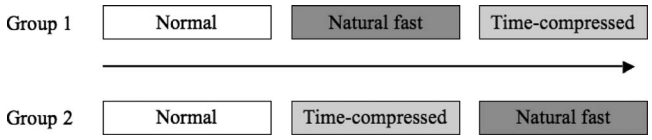


FIG. 1. Overview of the experimental design. Group 1 (top) was presented first with 60 normal sentences, immediately followed by 60 natural fast sentences, and followed by 60 time-compressed sentences. Group 2 (bottom) was presented first with 60 normal sentences, immediately followed by 60 time-compressed sentences, and followed by 60 natural fast sentences

generalized their newly learned ability to prevoiced alveolar syllables, while Tremblay *et al.* (1997) found a preattentive effect signaling transferred learning on listeners discriminating between prevoiced alveolar stops after having been trained to discriminate prevoiced labial stops. Finally, transfer of learning has been found for adaptation to a foreign accent across speakers (Bradlow and Bent, 2008). Bradlow and Bent (2008) found that listeners were better able to comprehend sentences in a foreign accent spoken by a novel speaker after having adapted to other speakers with the same foreign accent. In the present experiment, we tested whether having adapted to one type of fast speech facilitates adaptation and/or general performance for the other type. Time-compressed sentences differ from normal sentences only in their temporal characteristics, while natural fast sentences differ from normal sentences in their temporal characteristics as well as their spectral characteristics. Transfer of learning between the two speech types, and between temporal and spectral variations will be tested in the present experiment using a between-subjects design with two listener groups in which the order of presentation of the two speech types is varied. Both groups first verified 60 sentences spoken at a normal rate. During this normal-rate block, listeners could get used to the task and the type of sentences. Subsequently listeners in group 1 listeners verified 60 natural fast sentences before finally verifying 60 time-compressed sentences, while listeners in group 2 first verified 60 time-compressed sentences followed by 60 natural fast sentences. This division into two groups allowed us to study the effect of the type of compression (artificial or natural) on the adaptation process and to test whether there is transfer of learning. If there is transfer of learning from time-compressed speech to natural fast speech, then performance (i.e., accuracy) for the natural fast speech should be higher for group 2 than for group 1. Alternatively, if there is transfer of learning from natural fast speech to time-compressed speech, then performance for the time-compressed sentences should be higher for group 1 than for group 2. Figure 1 presents an overview of the order in which the three speech types were presented to both groups.

II. METHOD

A. Participants

Forty-two participants (nine male, mean age 22.1, stddev. 4.3 years, median age 22 years, range 18–41 years) took part in the study. All were native speakers of Dutch from The Netherlands, with no history of oral or written language impairment, or neurological or psychiatric disease. None reported any hearing problems or any previous experience with time-compressed speech. Listeners were randomly allocated to the two groups: 21 to group 1 and 21 to group 2. All gave written informed consent and were paid or received course credit for their participation.

B. Speech material

Recordings were made of a 31-year-old male speaker of Standard Dutch who had lived in The Netherlands all his life. Recordings were made of two versions of the 180 sentences listed in Table I. The procedure for the sentences produced at a normal rate was as follows. First, the sentence was presented on the computer screen in front of the speaker. He was instructed to first quietly read the sentence and to subsequently pronounce the sentence as a declarative statement at his normal speech rate. All sentences were recorded once. Next, the natural fast sentences were recorded. A sentence was presented on the computer screen. Again the speaker was asked to first read the sentence in silence. After that he produced the sentence four times in quick succession, as it was found that this was the best way for him to produce the sentences as fast and fluently as possible. The recordings were made in a sound-treated room, using a Sennheiser ME64 microphone, which was attached to an Alexis Multimix USB audio mixing station. The recordings were saved at 44 100 Hz to hard disk directly via an Imix DSP chip plugged into the Alexis Multimix and to the USB port of an Apple Macbook. PRAAT (Boersma and Weenink, 2003) was used to save all sentences into separate sound files with begin and end trimmed at zero crossings (trimming on or as closely as possible to the onset and offset of initial and final speech sounds) and resampled from 44 100 to 22 050 Hz. For the natural fast sentences, in the great majority of cases ($>95\%$), the second sentence was selected out of the quartet of sentences recorded, as these were judged by the experimenters to be the best examples (fastest as well as most fluent). Subsequently, the durations of the 2×180 sentences used in the experiment were calculated. The normal speech rate sentences consisted of 4.7 (intended) syllables on average (range 3–12 syllables, stdev. 0.6 syllables) and the speech rate of the natural fast sentences was 10.2 syllables/s (stddev. 1.6 syllables). On average, the selected natural fast sentences were pronounced at 46.0% of the duration of the normal speech rate sentences, with the fastest item pronounced at 32.6% and the slowest at 88.7%. Next, the time-compressed sentences were obtained by digitally shortening them with PSOLA (Pitch Synchronous Overlap and Add) (Moulines and Charpentier, 1990), as implemented in PRAAT. Compression rates were established per sentence: each individual time-compressed sentence was matched in rate to its corresponding natural fast item. For instance, if a natural fast sentence was pronounced at 48% of the duration of the normal speed sentence (i.e., twice as fast), then the compression rate for the PSOLA version of that sentence was set to 48%. Subsequently, the normal sentences and the natural fast sentences were all resynthesized at 100% of their original duration using PSOLA. Finally, the intensity of each of the 540 (180 sentences $\times$ 3 variants) sound files was peak normalized

at 99% of its maximum amplitude and scaled to 70 dB sound pressure level.

C. Procedure

All listeners were tested individually in a sound-treated booth and received written instructions. Responses were made using a button box with the index finger (true responses) and middle (false response) finger of their dominant hand. The stimuli were presented over Sennheiser HD477 headphones at a comfortable sound level per participant. Stimulus presentation and response time (RT) measurement were performed using PRESENTATION (Neurobehavioral Systems, Albany, CA). Response times were measured relative to the end of the audio file, following May *et al.* (2001) and Adank *et al.* (2009).

Each trial proceeded as follows. First, the stimulus sentence was presented. Second, the program waited for 3 s before playing the next stimulus, allowing the participant to respond. If the participant did not respond within 3 s, the trial was recorded as *no response*. Participants were asked to respond as quickly and accurately as possible and they were told that they did not have to wait until the sentence was finished (allowing for negative RTs, as RT was calculated from the offset of the sound file). Six familiarization trials were presented prior to the start of the experiment. The familiarization sentences had been produced by the same speaker and were spoken at a normal speech rate. The familiarization sentences were not included in the actual experiment. The test sentences were presented in a semirandomized order per participant and true and false sentences were counterbalanced across experimental blocks. Within an experimental condition, no true-false sentence pairs were presented. For instance, the true and false versions of sentence 2 ("Bevers bouwen dammen in de rivier" (English: *Beavers build dams in the river*) and "Bevers groeien in een moestuin" (English: *Beavers grow in the vegetable patch*), see Table I), were never presented within one experimental condition. Total duration of the listening study was 15 min, without breaks.

III. RESULTS

The data from one of the participants of group 1 were excluded from the analysis, as her average RTs were more than two standard deviations slower than the average across all participants. Due to a programming error, six participants (three per listener group) got 70 (instead of 60) time-compressed sentences and they then got 50 (instead of 60) natural fast sentences. We excluded the last ten time-compressed trials for these participants and recoded trial number within the natural fast block of sentences.

Figure 2 and Table II show the average error percentages for both groups per speech type for the data grouped into ten subsequent miniblocks of sentences, in order to see adaptation over exposure time. Likewise, Fig. 3 and Table III show average RTs for the two groups (in milliseconds, measured from sentence offset) for the three speech types, again broken down in ten (mini)blocks of six sentences. The results in Figs. 2 and 3 are only plotted in ten miniblocks of six subsequent sentences for demonstration purposes. The statistical analysis was performed with trial as a continuous variable.

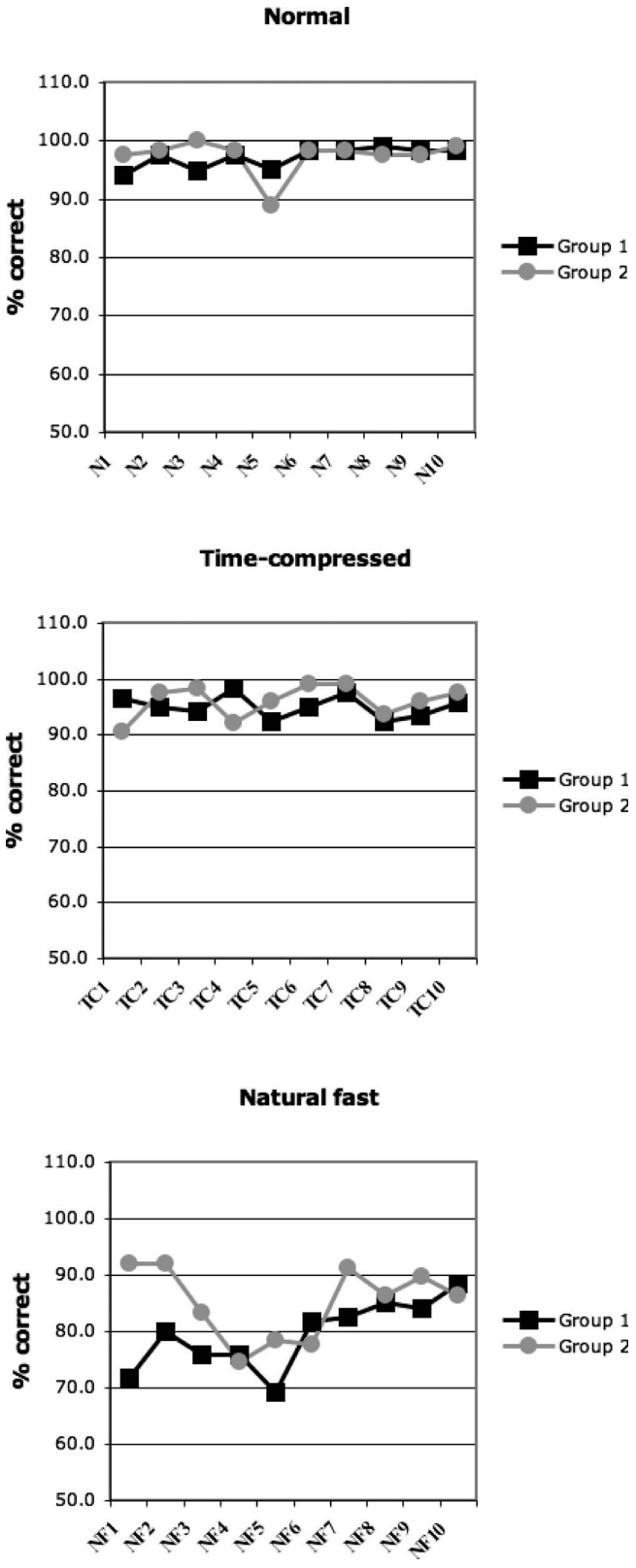


FIG. 2. Average percent correct (%) per miniblock of six sentences correct for the normal speed condition (top panel, miniblocks N1–N10), Time-compressed condition (middle panel, miniblocks TC1–TC10), and the natural fast condition 1 (bottom panel, miniblocks NF1–NF10) for both groups (group 1 in black and group 2 in gray).

The results were analyzed with linear mixed effects models with participant and item as crossed random effects

TABLE II. Mean RTs in ms plus standard deviations (Stddev) for both groups for the three speech types for the ten blocks of six sentences.

RT (ms)		Normal		Time compressed		Natural fast	
		Mean	Stddev	Mean	Stddev	Mean	Stddev
Group 1	Block 1	231	304	676	409	918	528
	Block 2	261	395	491	301	863	508
	Block 3	280	290	474	303	800	484
	Block 4	223	266	482	312	799	442
	Block 5	264	296	496	340	876	502
	Block 6	223	341	471	383	800	480
	Block 7	257	316	524	413	695	371
	Block 8	292	387	496	334	762	453
	Block 9	266	333	490	302	771	489
	Block 10	232	299	485	316	796	497
Group 2	Block 1	231	304	565	381	745	493
	Block 2	261	395	481	346	762	516
	Block 3	280	290	520	365	791	477
	Block 4	223	266	453	278	832	546
	Block 5	264	296	443	288	727	440
	Block 6	223	341	482	341	767	471
	Block 7	257	316	478	273	703	446
	Block 8	292	387	495	303	809	556
	Block 9	266	333	501	308	744	460
	Block 10	232	299	446	335	818	513

(Pinheiro and Bates, 2000; Quené and van den Bergh, 2004). One model was fitted to the binomial accuracy data (a response being correct or incorrect), and one model was fitted to the RT data (for correct responses only). Order was a between-participant factor, and speech type (normal, time compressed, or natural fast) and trial (within each block of 60 sentences of that particular speech type) were within-participant factors. As mentioned above, we chose to look for effects of trial to study adaptation, rather than of miniblock (see Figs. 2 and 3) because trial provided us with the most fine-grained continuous variable in relation to adaptation (note, though, that an alternative analysis with the variable miniblock, instead of trial, produced highly similar results). We also entered the (within-participants and between-items) factor of whether the sentence ought to elicit a true or a false response because participants may have found it easier to verify either type. Systematic stepwise model comparisons using likelihood ratio tests established the best-fitting model.

A. Accuracy

The linear mixed-effects model for accuracy had as dependent variable whether or not the response was correct ($N=7320$). The within-subjects factor speech type had three levels (normal, time compressed, and natural fast). The linear mixed effects model gives as output whether each of the levels differs significantly from the one mapped onto the intercept (in this case, the normal-rate sentences). Beta values are provided for significant effects and interactions (with standard error in brackets) as well as significance levels.

Performance on the natural-fast sentences was significantly poorer than performance on the normal-rate sentences [$\beta=-2.234$ (0.382), $p<0.001$], but performance on the time-compressed sentences was not. There was an overall effect of trial [$\beta=0.040$ (0.014), $p<0.01$], indicating that performance improved over trials within speech type. The effect of correct response (true or false) also significantly affected accuracy: participants showed better performance for the false than the true sentences [$\beta=0.708$ (0.236), $p<0.01$]. Overall, the two listener groups did not differ in performance [order: $\beta=1.048$ (0.552), n.s.].

Speech type interacted with trial: for the time-compressed speech, improvement over trials was less than in the other two speech types [$\beta=-0.041$ (0.015), $p<0.05$]. This was modified further by a three-way interaction of order by speech type by trial [$\beta=0.0584$ (0.022), $p<0.01$]: that there was less improvement over trials for the time-compressed speech, relative to the other speech types, was mainly the case for the listeners in group 1, who heard the time-compressed sentences after they had been presented with the natural-fast speech. It was less true for the group 2 listeners who heard the time-compressed sentences before the natural-fast sentences. This fits in with slightly poorer overall performance for group 2 on the time-compressed sentences, as suggested by an order by speech type interaction [$\beta=-1.469$ (0.707), $p<0.05$] for the time-compressed speech.

The data were also analyzed for the three speech types separately to investigate whether there is improvement or adaptation over trials and to see whether the order in which listeners heard the conditions mattered. Bonferroni correction was applied to the outcomes of the subset analyses (we analyzed three subsets and the critical p-value of 0.05 was thus set to $0.05/3$, resulting in a critical value of 0.017).

For the normal-rate sentences, there was an overall ef-

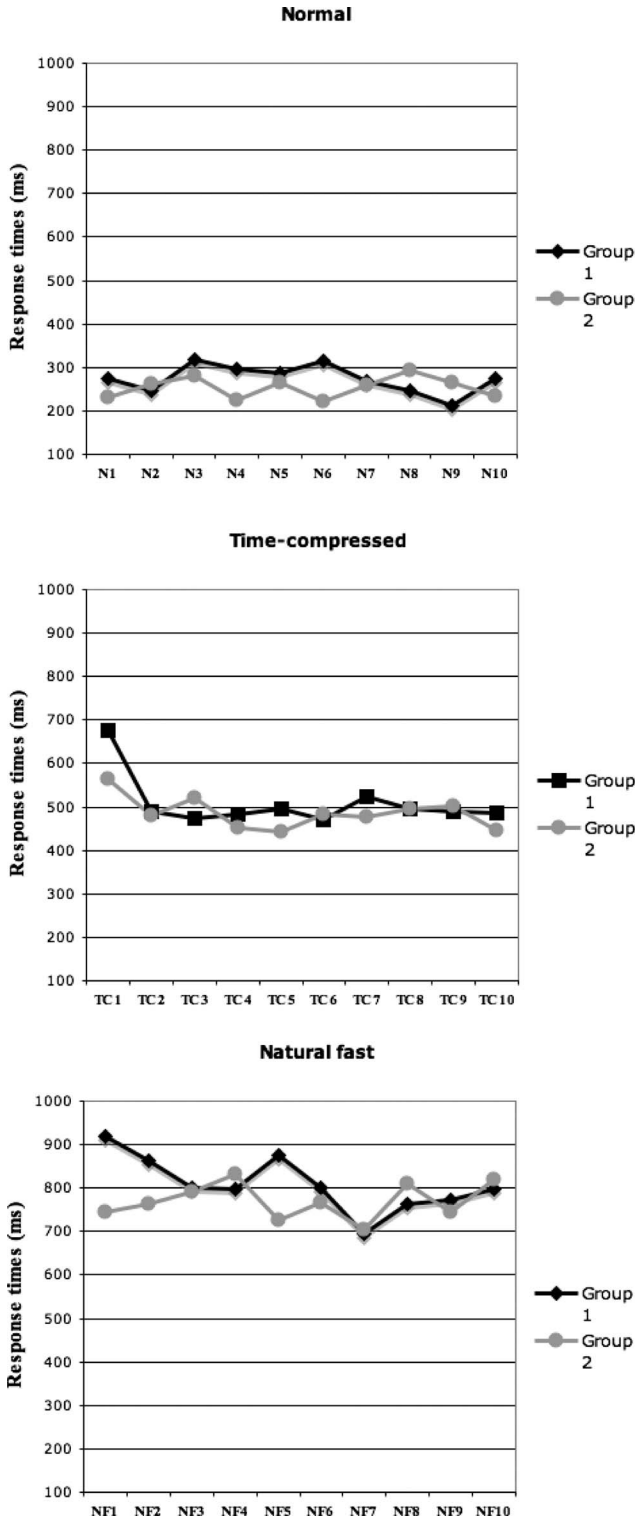


FIG. 3. Average RTs in millisecond per miniblock of six sentences correct for the normal speed condition (top panel, miniblocks N1–N10), time-compressed condition (middle panel, miniblocks TC1–TC10), and the natural fast condition 1 (bottom panel, miniblocks NF1–NF10) for both groups (group 1 in black and group 2 in gray).

fect of trial, meaning that accuracy performance improved over trials [β=0.050 (0.025), p=0.044], but note that this does not exceed the Bonferroni-corrected critical value for significance. There was no difference between the two orders (i.e., between the two listener groups, and note that normal-rate sentences were presented first in both orders) and no interaction between trial and order. The effect of correct response (true or false) was not significant in this subset.

For the time-compressed sentences, there was no overall effect of trial and no interaction between order and trial. The only effect approaching significance was that of correct response: stating that the sentence is false being the easier response [β=1.091 (0.483), p=0.024, which does not meet the Bonferroni-corrected threshold value]. This subanalysis complements the picture provided by the two-way and three-way interactions reported above in the overall analysis. Unlike the other speech conditions, there is no improvement in accuracy over time-compressed trials (this was particularly the case if the time-compressed sentences were presented as the last speech condition, but when the time-compressed condition preceded the natural fast condition improvement over trials was not significant either).

For the natural fast sentences, there was an overall order effect [β=0.957, (0.29), $p<0.001$]. This shows that listeners who got this condition last (i.e., after they had been presented with the time-compressed condition) had overall higher accuracy than listeners who got this condition before the time-compressed condition. Second, there was an overall effect of trial [β=0.022 (0.008), $p<0.01$], indicating that accuracy improved over trials. Furthermore, there was an order by trial interaction [β=−0.016 (0.008), p=0.047], showing that listeners who got the natural fast sentences last showed a smaller improvement over trials than listeners who got the natural fast sentences before the time-compressed sentences (note though that this interaction fails to reach significance if we take the Bonferroni correction into account). Finally, there was a significant effect of correct response, which means that false sentences were easier to verify than true sentences [β=0.561 (0.275), $p<0.05$]. This subset analysis clearly shows that both listener groups showed improvement over the course of the 60 natural fast sentences and that order mattered: the group who had already been presented with the time-compressed materials had overall better performance than the other group.

B. Response times

Figure 3 and Table III show the average RTs per speech type. The results are again plotted in ten (mini)blocks of six subsequent sentences each. The statistical analysis, as in the accuracy analysis, was performed with trial as a continuous variable. A linear mixed effect model was fitted to the RTs (measured from sentence offset) of the correct decisions (N=6716). As in the previous analysis, the linear mixed effect model gives as output whether each of the levels differs significantly from the one mapped onto the intercept (i.e., the normal-rate sentences).

Response times were significantly longer in the time-compressed condition than in the normal-rate condition [β=256 (21.5), $p<0.001$]. The same was true for the natural fast sentences [β=452 (22.7), $p<0.001$]: RTs were longer compared to the normal-rate sentences. There were no overall effects of order, trial, or correct response. Correct response did interact with speech type, however: in the natural fast sentence condition, listeners took longer to decide that sentences were false [β=239 (19.9), $p<0.001$]. Even though

there was no overall trial effect, there were significant interactions between speech type and trial. In the time-compressed condition, the effect of trial differed from that in the normal-rate condition [$\beta=-1.164$ (0.547), $p<0.05$], suggesting that responses did get faster over the time-compressed trials. In the natural fast condition, the trial effect was also different from that in the normal-rate condition [$\beta=-1.450$ (0.576), $p<0.05$], suggesting that responses did get faster over the natural fast trials. None of the other interactions proved significant.

As in the accuracy analysis, RTs were also analyzed per speech condition to complement the picture of the overall analysis. Bonferroni correction was applied to the critical value for these subset analyses ($0.05/3=0.017$). For the normal-rate sentences, there were no significant effects of order, trial, or of correct response. There were no significant interactions either. For the time-compressed speech, there was a significant effect of trial [$\beta=-1.534$ (0.641), $p=0.017$, which just satisfies the Bonferroni corrected critical value]. Figure 3 shows that this speeding up of responses over trials was found mainly in the initial two-three miniblocks. There was no effect of order or of correct response. The interaction between trial and order was not significant either, indicating that listeners in both order groups got faster over trials. For the natural fast sentences, the data showed an effect of trial [$\beta=-2.091$ (1.111), $p=0.060$, which does not meet the criterion for significance] and of correct response [$\beta=214.9$ (40.79), $p<0.001$]. There was no interaction between order and trial, which means that both groups tended to become somewhat faster over trials.

The results for the time-compressed speech replicate results from Clarke and Garrett (2004), who found that listeners got faster at a RT task after presentation of a small number of sentences. Our results show that group 1 got 185 ms faster between the first and the second miniblock of six sentences, while group 2 became 84 ms faster.

In sum, the RT analysis clearly confirms the difficulty hierarchy of the three speech types also seen in the accuracy scores: listeners were fastest to respond to the normal-speed sentences, slower for the time-compressed sentences, and slowest for the natural fast sentences. The RTs were not affected by the order in which the two fast speech types were presented. Importantly, whereas adaptation to time-compressed speech did not show up as improved accuracy over trials, it was found in decreased RTs over trials. Adjustment to natural fast speech was found both in improved accuracy and in somewhat decreased RTs over trials.

Finally, one should note that any learning observed in the normal-rate condition indicates that participants needed more sentences than the six sentences in the familiarization block to get used to the task of sentence verification. Even if accuracy over the first half (30) of the normal-rate sentences is compared to accuracy in the second half, performance is significantly better in the second half. The importance of ruling out rival explanations (such as practice effects) for improved performance over trials has always been an issue in adaptation studies (Clarke and Garrett, 2004; Dupoux and Green, 1997).

IV. GENERAL DISCUSSION

We sought to establish whether listeners learn to adapt to naturally fast speech and if so, how this process compares to learning to adapt to time-compressed speech. Two groups of listeners participated in a speeded sentence verification experiment. Both groups first verified a series of sentences at a normal speaking rate. Subsequently, listeners in group 1 verified a series of natural fast sentences, followed by a series of time-compressed sentences, while this order was reversed for group 2.

The results have shown three important points. First, listeners adapt to natural fast speech. Gradual adaptation had been shown for artificially time-compressed speech materials, but not yet for natural fast speech. Natural fast speech involves a greater spectrotemporal deviation from a normal-rate speech signal than artificial time compression. Listeners' performance clearly showed that natural-fast speech is more difficult to process than artificially time-compressed speech due to the greater spectrotemporal variation, as was previously shown in Janse, 2004. The present finding that listeners are nevertheless able to adapt to natural fast speech complements the earlier findings of adaptation to highly compressed speech.

The second important point is that we have shown transfer of learning from adaptation to time-compressed speech to naturally produced fast speech. The group who had been presented with time-compressed material *before* they were presented with the natural fast material (group 2) showed generally higher accuracy for the natural fast materials. Listeners in this group benefited from having already adapted to the temporal manipulation—the time-compressed sentences—before being presented with sentences that showed temporal compression as well as spectral variation. Furthermore, their adaptation curve was shallower, because they started off higher, than that of the group who got the natural fast sentences first.

Third, whether there is transfer of learning from the natural fast speech to the time-compressed condition was less clear. One could argue that if listeners had adapted to natural fast speech, which involves a fast rate and greater spectral smearing, time-compressed speech ought to be relatively easy. Our results do not confirm this argument, however. Both groups showed adaptation to artificial time compression in terms of decreased RTs over trials and there was no evidence for a difference in slope. Apparently, transfer of learning shows up more clearly if one is presented with speech conditions of increasing complexity rather than if the most difficult condition is followed by an easier condition.

The present study replicated the effect of learning on reaction times (Clarke and Garrett, 2004) for the time-compressed speech. Participants became faster but not more accurate for the time-compressed sentences. However, they became more accurate and somewhat faster for the natural fast sentences. This difference between time-compressed and natural fast speech may be explained by the overall difficulty of the two speech types: listeners in both groups made more errors and showed longer RTs for the natural fast sentences than for the time-compressed sentences. After presentation of

approximately 30 sentences, they were able to understand the sentences better, but they still needed longer processing to perform the task adequately.

In the experiment, the time-compression factor varied per stimulus. The sentences in the time-compressed condition were matched in compression factor with the natural fast sentences. It is unclear how this may have affected the extent to which participants adapted to the manipulation. There is some evidence that phonetic variability during exposure/training aids perceptual learning (Logan *et al.*, 1991). However, one study on adapting to time-compressed speech shows that a change in compression rate can lead to a temporary *decrease* in performance (Dupoux and Green, 1997), while another study shows that a change in compression rate does not affect performance (Golomb *et al.*, 2007). The initial decrease in RT for the time-compressed condition (see Fig. 3) seems to be in line with Golomb *et al.* (2007) that even continuous changes in compression rate did not hinder adaptation to time-compressed speech.

In sum, our results show that listeners adapted to time-compressed speech and natural fast speech and that there was a transfer of learned skills from time-compressed to natural fast sentences, but not the other way around. Adapting to time-compressed speech has been studied extensively in the past decades, and several explanations have been suggested. For instance, adaptation to time-compressed speech has often been described as an attention-weighing process in which listeners shift their attention from task-irrelevant to task-relevant cues (Goldstone, 1998; Golomb *et al.*, 2007; Nosofsky, 1986). Moreover, it has been argued that learning of time-compressed speech is characterized by the recalibration of the boundaries between speech sounds to accommodate the faster speech rate (Golomb *et al.*, 2007). In the discussion below, we attempt to further elucidate the type of cognitive processing underlying adaptation, using Ahissar and Hochstein's (2004) reverse hierarchy theory (RHT), a theory for perceptual learning and transfer (see also Amitay, 2009).

In RHT, perceptual learning is defined as practice-induced improvements in the ability to perform specific perceptual tasks. These improvements involve explicit and extensive practice, for instance, when learning to understand a new language. RHT poses that perceptual learning stems largely from a gradual top-down processing cascade during which first higher and then lower-level task-relevant cues become available. During this process, task-relevant cues are enhanced and task-irrelevant cues are filtered out.

RHT makes explicit predictions about the role of attention and task difficulty on processing level and transfer of learning. With respect to the level of processing, RHT predicts that the cascade from high to low levels of processing is top down and guided by attention as task difficulty increases. When difficulty increases, attention becomes more focused to lower processing levels and lower-level cues become more relevant for task improvement. When applied to our data, this prediction implies that participants relied more on lower-level acoustic cues for conditions that required more attention, i.e., those that were more difficult. It seems plausible that the natural fast condition was the most difficult condition in the experiment as performance was less accurate and slower. Following RHT's prediction, this implies that perceptual learning for the natural fast condition relied more on lower-level acoustic cues than learning of the time-compressed condition. Recall that participants had to process variation resulting from the applied temporal compression while adapting to time-compressed sentences, while for the natural fast sentences they had to adapt to temporal compression *and* to spectral variability. For the natural fast sentences, RHT thus predicts that the higher difficulty of the natural fast sentences condition led them to direct their attention more to lower-level (possibly spectral) acoustic cues than was the case in the time-compressed sentence condition. Further studies are required to address the speculation that spectral and temporal variabilities may be dealt with at different processing levels.

With respect to transfer of learning, RHT predicts that learning at higher processing levels results in more transfer, while learning at lower levels leads to more specificity. RHT also predicts that task difficulty of a preceding task affects learning in the subsequent task. Transfer of learning occurs when an easy condition is followed by a more difficult task, but not when a difficult task is followed by an easier task (Ahissar and Hochstein, 1997; Liu *et al.*, 2008; Pavlovskaya and Hochstein, 2004). Our results comply with this prediction, as we observed task improvement for the natural fast condition (the more difficult task) when it was preceded by the time-compressed condition (the easier task), but not when the natural fast condition preceded the time-compressed condition. Ahissar and Hochstein (2004) suggested that training on easier tasks enables lower-level learning associated with difficult tasks. This suggests for our data that adapting to time-compressed condition, which may involve learning at higher processing levels, improved performance in the natural fast condition by enabling the focus of attention on lower-level cues. As said, learning at lower levels would then lead to more specificity and less transfer from the natural fast to the time-compressed condition.

In conclusion, our results have shown that listeners adapt to extremely fast naturally produced speech. This result is highly relevant because it complements previous research of the learnability of artificially time-compressed speech. Finally, the present results provide one further demonstration of the flexibility of the human speech comprehension system and its ability to adapt on-line to novel variation sources in the speech signal. Our results thus add to a growing body of research on adaptation to natural and artificial variations in the speech signal.

ACKNOWLEDGMENTS

We wish to thank Erik van den Boogert for technical assistance and Matthijs Noordzij for lending his voice. This research was supported by the Netherlands Organization for Research (NWO) under Project Nos. 275-75-003 (P.A.) and 275-75-004 (E.J.).

APPENDIX

See Table III.

TABLE III. True and false Dutch sentences used in the experiment.

No.	True	False
1	Makrelen ademen door kieuwen	Chirurgen groeien aan planten
2	Bevers bouwen dammen in de rivier	Bevers groeien in een moestuin
3	Bisschoppen dragen kleren	Wortels hebben een beroep
4	Ezels dragen zware vrachten	Bromfietsen hebben een snavel
5	Pinguins eten veel vis	Slagers hebben een staart
6	Tomaten groeien aan planten	Forellen hebben een vacht
7	Wortels groeien in een moestuin	Haaien hebben handen
8	Architecten hebben een beroep	Nachtegalen hebben manen
9	Roodborstjes hebben een snavel	Pinguns hebben schubben
10	Tijgers hebben een staart	Tomaten hebben sterke tanden
11	Luipaarden hebben een vacht	Makrelen hebben veren
12	Vaders hebben handen	Lepels hebben vier poten
13	Leeuwen hebben manen	Schuurtjes hebben voelsprieten
14	Forellen hebben schubben	Aardappels hebben voeten
15	Haaien hebben sterke tanden	Leeuwen hebben winkels
16	Nachtegalen hebben veren	Mieren zijn van hout
17	Beren hebben vier poten	Vlinders komen van schapen
18	Vlinders hebben voelsprieten	Hamers kruipen op hun buik
19	Wetenschappers hebben voeten	Auto's kunnen goed zwemmen
20	Slagers hebben winkels	Tantes kunnen in winkels gekocht worden
21	Kasten zijn van hout	Kroketten kunnen koppig zijn
22	Lammetjes komen van schapen	Asperges kunnen ver vliegen
23	Ratelslangen kruipen op hun buik	Messen zijn eetbaar
24	Otters kunnen goed zwemmen	Biefstukken moeten lang studeren
25	Blikopeners kunnen in winkels gekocht worden	Wijnflessen rijden op de weg
26	Ezels kunnen koppig zijn	Wandelschoenen vliegen rond op zoek naar voedsel
27	Ganzen kunnen ver vliegen	Luipaarden voeren het bevel op scheppen
28	Druiven zijn eetbaar	Roodborstjes werken in de politiek
29	Chirurgen moeten lang studeren	Ezels wonen in een klooster
30	Bromfietsen rijden op de weg	Ratelslangen worden gebruikt als keukengerei
31	Bijen vliegen rond op zoek naar voedsel	Presidenten worden gebruikt voor het eten van soep
32	Kapiteins voeren het bevel op schepen	Kapiteins worden gebruikt voor opslag
33	Presidenten werken in de politiek	Monniken worden geschild
34	Monniken wonen in een klooster	Tijgers worden gemaakt in een fabriek
35	Messen worden gebruikt als keukengerei	Taarten worden in de tuin gebruikt
36	Lepels worden gebruikt voor het eten van soep	Architecten worden verkocht door slagers
37	Schuurtjes worden gebruikt voor opslag	Politieagenten hebben een kurk
38	Aardappels worden geschild	Heggenscharen zijn altijd vrouwen
39	Sloffen worden gemaakt in een fabriek	Ezels zijn deel van de familie
40	Heggenscharen worden in de tuin gebruikt	Giraffes zijn fruit
41	Biefstukken worden verkocht door slagers	Wetenschappers zijn gefabriceerde goederen
42	Wijnflessen hebben een kurk	Beren zijn gefrituurd
43	Tantes zijn altijd vrouwen	Ganzen zijn groenten
44	Ooms zijn deel van de familie	Ministers worden in een oven gebakken
45	Bananen zijn fruit	Olifanten zijn klein
46	Wandelschoenen zijn gefabriceerde goederen	Kasten zijn levende wezens
47	Kroketten zijn gefrituurd	Kakkerlakken zijn meubels
48	Asperges zijn groenten	Ooms zijn om op te zitten
49	Taarten worden in een oven gebakken	Dolfijnen gebruiken benzine
50	Mieren zijn klein	Sloffen zijn insecten
51	Olifanten zijn levende wezens	Bananen zijn zoogdieren
52	Tafels zijn meubels	Vaders zitten in de gereedschapskist
53	Stoelen zijn om op te zitten	Lammetjes zitten in de regering
54	Auto's gebruiken benzine	Bijen hebben een lange nek
55	Kakkerlakken zijn insecten	Stoelen lopen op straat
56	Dolfijnen zijn zoogdieren	Een kameel is een soort vogel
57	Hamers zitten in de gereedschapskist	Een panter heeft vleugels
58	Ministers zitten in de regering	Een kool is een soort vrucht
59	Giraffes hebben een lange nek	Een boon is zoet
60	Politieagenten lopen op straat	Een mus is een zoogdier
61	Een pelikaan is een soort vogel	Een overhemd is een lichaamsdeel
62	Een adelaar heeft vleugels	Een schoen heeft vingers
63	Een aardbei is een soort vrucht	Een aap is een soort vis
64	Een appel is zoet	Een boor is een muziekinstrument
65	Een varken is een zoogdier	Een viool is een werktuig
66	Een been is een lichaamsdeel	Mensen dragen een broek aan hun handen

TABLE III. (*Continued.*)

No.	True	False
67	Een hand heeft vingers	Sommige mensen hebben giraffes als huisdier
68	Een stekelbaars is een soort vis	De meeste auto's rijden op appelsap
69	Een gitaar is een muziekinstrument	Denemarken is een land in Afrika
70	Een waterpomptang is gereedschap	Een paard heeft drie benen
71	Mensen dragen sokken aan hun voeten	Roken is goed voor je gezondheid
72	Sommige mensen hebben honden als huisdier	Een uur is vijfenveertig minuten
73	De meeste vrachtwagens rijden op diesel	Melk bevat alcohol
74	Spanje is een land in Europa	Mensen hebben op de zon gelopen
75	Een paard heeft vier benen	Sommige mensen drinken thee met zout
76	Beweging is goed voor je gezondheid	Olifanten eten soms mensen op
77	Een minuut heeft zestig seconden	Een boom heeft melk nodig om te leven
78	Bier bevat alcohol	Een groen licht betekent stop
79	Mensen hebben op de maan gelopen	Papier wordt gemaakt van onkruid
80	Sommige mensen drinken koffie met suiker	Een fiets is een oorlogwapen
81	Krokodillen eten soms kinderen op	Boeddhisme is een politieke theorie
82	Een plant heeft water nodig om te leven	Spaghetti is een Frans gerecht
83	Een rood licht betekent stop	Een loodgieter kan je helpen als je ziek bent
84	Perkament wordt gemaakt van leer	Fietsen is meestal langzamer dan lopen
85	Een tank is een oorlogwapen	Kinderen zijn nooit bang in het donker
86	Baksteen is een goed materiaal voor gebouwen	Een schip is een soort meubel
87	Boekhouden is een beroep	Een sinaasappel is knapperig
88	Juni is een zomermaand	Een baksteen is een edelsteen
89	Een step is goed te besturen	De hoofdstad van Nederland is Brussel
90	Een vrachtwagen heeft een motor	Een kip kan goed gitaar spelen

Adank, P., Evans, B. G., Stuart-Smith, J., and Scott, S. K. (**2009**). "Comprehension of familiar and unfamiliar native accents under adverse listening conditions," J. Exp. Psychol. Human **35**, 520–529.

Ahissar, M., and Hochstein, S. (**1997**). "Task difficulty and the specificity of perceptual learning," Nature (London) **387**, 401–406.

Ahissar, M., and Hochstein, S. (**2004**). "The reverse hierarchy of visual perceptual learning," Trends Cogn. Sci. **8**, 457–464.

Amitay, S. (**2009**). "Forward and reserve hierarchies in auditory perceptual learning," Learn. Perception **1**, 59–68.

Baddeley, A. D., Emslie, H., and Nimmo-Smith, I. (**1992**). "The speed and capacity of language processing (SCOLP) test," Bury St Edmunds: Thames Valley Test Company.

Boersma, P., and Weenink, D. (**2003**). "Praat: Doing phonetics by computer," http://www.praat.org (Last viewed 10/28/2006).

Bradlow, A. R., and Bent, T. (**2008**). "Perceptual adaptation to non-native speech," Cognition **106**, 707–729.

Browman, C., and Goldstein, L. (**1990**). "Tiers in articulatory phonology, with some implications for casual speech," in *Papers in Laboratory Phonology I*, edited by J. Kingston and M. E. Beckman (Cambridge University Press, Cambridge), pp. 341–376.

Byrd, D., and Tan, C. C. (**1996**). "Saying consonant clusters quickly," J. Phonetics **24**, 263–282.

Clarke, C. M., and Garrett, M. F. (**2004**). "Rapid adaptation to foreign-accented English," J. Acoust. Soc. Am. **116**, 3647–3658.

Cormier, S. M., and Hagman, J. D. (**1987**). *Transfer of Learning: Contemporary Research and Applications*. (Academic, San Diego).

Delhommeau, K., Micheyl, C., and Jouvent, R. (**2005**). "Generalization of frequency discrimination learning across frequencies and ears: Implications for underlying neural mechanisms in humans," J. Assoc. Res. Otolaryngol. **6**, 171–179.

Delhommeau, K., Micheyl, C., Jouvent, R., and Collet, L. (**2002**). "Transfer of learning across durations and across ears in auditory frequency discrimination," Percept. Psychophys. **64**, 426–436.

Dupoux, E., and Green, K. (**1997**). "Perceptual adjustment to highly compressed speech: Effects of talker and rate changes," Immunopharmacol Immunotoxicol **23**, 914–927.

Ernestus, M., Baayen, H., and Schreuder, R. (**2002**). "The recognition of reduced forms," Brain Lang **81**, 162–173.

Goldstone, R. L. (**1998**). "Perceptual learning," Annu. Rev. Psychol. **49**, 585–612.

Golomb, J., Peelle, J. E., and Wingfield, A. (**2007**). "Effects of stimulus variability and adult aging on adaptation to time-compressed speech," J. Acoust. Soc. Am. **121**, 1701–1708.

Green, K. P., Stevens, K. N., and Kuhl, P. K. (**1994**). "Talker continuity and the use of rate information during phonetic perception," Percept. Psychophys. **55**, 249–260.

Haskell, R. E. (**2001**). *Transfer of Learning: Cognition, Instruction and Reasoning* (Academic, San Diego).

Janse, E. (**2004**). "Word perception in fast speech: Artificially time-compressed vs. naturally produced fast speech," Speech Commun. **42**, 155–173.

Koreman, J. (**2006**). "Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech," J. Acoust. Soc. Am. **119**, 582–596.

Lehiste, I. (**1970**). *Suprasegmentals* (MIT, Cambridge, MA).

Liu, E. H., Mercado, E., Church, B. A., and Orduna, I. (**2008**). "The easy-to-hard effect in human (Homo Sapiens) and rat (Rattus Norvegicus) auditory identification," J. Comp. Psychol. **122**, 132–145.

Logan, J. S., Lively, S. E., and Pisoni, D. (**1991**). "Training Japanese listeners to identify English /r/ and /l/: A first report," J. Acoust. Soc. Am. **89**, 874–886.

Max, L., and Caruso, A. J. (**1997**). "Acoustic measures of temporal intervals across speaking rates: Variability of syllable- and phrase-level relative timing," J. Speech Lang. Hear. Res. **40**, 1097–1110.

May, J., Alcock, K. J., Robinson, L., and Mwita, C. (**2001**). "A computerized test of speed of language comprehension unconfounded by literacy," Appl. Cognit. Psychol. **15**, 433–443.

McClaskey, C., Pisoni, D., and Carrell, T. (**1983**). "Transfer of learning of a new linguistic contrast in voicing," Percept. Psychophys. **34**(4), 323–330.

Miller, J. L., and Liberman, A. M. (**1979**). "Some effects of later-occurring information on the perception of stop consonant and semivowel," Percept. Psychophys. **25**, 457–465.

Miller, J. L., Aibel, I. L., and Green, K. P. (**1984a**). "On the nature of rate-dependent processing during phonetic perception," Percept. Psychophys. **35**, 5–15.

Miller, J. L., Grosjean, F., and Lomanto, C. (**1984b**). "Articulation rate and its variability in spontaneous speech: A reanalysis and some implication," Phonetica **41**, 215–225.

Moulines, E., and Charpentier, F. (**1990**). "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," Speech Commun. **9**, 453–467.

Nosofsky, R. M. (**1986**). "Attention similarity, and the identification-specific relationship," J. Exp. Psychol. Gen. **115**, 39–57.

Pallier, C., Sebastián-Gallés, N., Dupoux, E., Christophe, A., and Mehler, J. (**1998**). "Perceptual adjustment to time-compressed speech: A cross-linguistic study," Mem. Cognit. **26**, 844–851.

Pavlovskaya, M., and Hochstein, S. (**2004**). "Transfer of perceptual learning effects to untrained stimulus dimensions," J. Vision **4**, 416.
Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**, 693–703.
Pinheiro, J., and Bates, D. (**2000**). *Mixed Effects Models in S and S-Plus* (Springer, New York).
Quené, H., and Van den Bergh, H. (**2004**). "On multi-level modeling of data from repeated measures designs: A tutorial," Speech Commun. **43**, 103–121.
Sebastián-Gallés, N., Dupoux, E., Costa, A., and Mehler, J. (**2000**). "Adaptation to time-compressed speech: Phonological determinants," Percept. Psychophys. **62**, 834–842.
Thorndike, E. L., and Woodforth, R. S. (**1901**). "The influence of the improvement in one mental function upon the efficiency of other functions. I.," Psychol. Rev. **8**, 553–656.
Tremblay, K., Kraus, N., Carrell, T. D., and McGee, T. (**1997**). "Central auditory system plasticity: Generalization to novel stimuli following listening training," J. Acoust. Soc. Am. **102**, 3762–3773.
Wingfield, A., Peelle, J. E., and Grossman, M. (**2003**). "Speech rate and syntactic complexity as multiplicative factors in speech comprehension by young and older adults," Aging Neuropsychol. Cogn. **10**, 310–322.
Wouters, J., and Macon, M. W. (**2002**). "Effects of prosodic factors on spectral dynamics. I. Analysis," J. Acoust. Soc. Am. **111**, 417–427.

Perceptual adaptation and intelligibility of multiple talkers for two types of degraded speech

Tessa Bent[a)]
Department of Speech and Hearing Sciences, Indiana University, 200 South Jordan Avenue, Bloomington, Indiana 47405

Adam Buchwald
Department of Speech-Language Pathology and Audiology, New York University, 665 Broadway, Suite 910, New York, New York 10012

David B. Pisoni
Department of Psychological and Brain Sciences, Indiana University, 1101 East 10th Street, Bloomington, Indiana 47405

(Received 7 July 2008; revised 29 July 2009; accepted 31 July 2009)

Talker intelligibility and perceptual adaptation under cochlear implant (CI)-simulation and speech in multi-talker babble were compared. The stimuli consisted of 100 sentences produced by 20 native English talkers. The sentences were processed to simulate listening with an eight-channel CI or were mixed with multi-talker babble. Stimuli were presented to 400 listeners in a sentence transcription task (200 listeners in each condition). Perceptual adaptation was measured for each talker by comparing intelligibility in the first 20 sentences of the experiment to intelligibility in the last 20 sentences. Perceptual adaptation patterns were also compared across the two degradation conditions by comparing performance in blocks of ten sentences. The most intelligible talkers under CI-simulation also tended to be the most intelligible talkers in multi-talker babble. Furthermore, listeners demonstrated a greater degree of perceptual adaptation in the CI-simulation condition compared to the multi-talker babble condition although the extent of adaptation varied widely across talkers. Listeners reached asymptote later in the experiment in the CI-simulation condition compared with the multi-talker babble condition. Overall, these two forms of degradation did not differ in their effect on talker intelligibility, although they did result in differences in the amount and time-course of perceptual adaptation.
[DOI: 10.1121/1.3212930]

PACS number(s): 43.71.Gv [RSN] Pages: 2660–2669

I. INTRODUCTION

Although it is well known that talkers differ in intelligibility (Bond and Moore, 1994; Bradlow *et al.*, 1996; Hazan and Markham, 2004; Hood and Poole, 1980), less is known about the stability of these differences across different types of signal degradation. In this paper, we ask whether the talkers who are highly intelligible in multi-talker babble are also highly intelligible when their speech is processed by a cochlear implant (CI) simulator. In addition, we investigate whether some talkers are easier or harder to adapt to. Finally, we report on how the type of degradation contributes to the process of adaptation.

A. Speech intelligibility

What factors determine speech intelligibility?[1] Broadly speaking, traditional views of speech intelligibility have maintained that intelligibility is a property of the speaker, the acoustic signal, or of the specific words being perceived (Black, 1957; Bond and Moore, 1994; Bradlow *et al.*, 1996; Hood and Poole, 1980; Howes, 1952, 1957). There is empirical evidence supporting each of these views. For example, certain properties of words (e.g., segmental composition, length, and frequency) have been shown to influence speech intelligibility (Black, 1957; Howes, 1952, 1957). Similarly, it has been shown that several specific acoustic properties of a talker's speech (e.g., speaking rate and vowel dispersion) play a crucial role in determining speech intelligibility (Bond and Moore, 1994; Bradlow *et al.*, 1996; Hood and Poole, 1980), as can properties of the listening environment (e.g., Assman and Summerfield, 2004; Fletcher and Steinberg, 1924; Miller, 1947; Miller and Nicely, 1955). However, the speech materials and the talker are not the only relevant factors in determining speech intelligibility. Instead, a variety of research findings suggest that speech intelligibility is also influenced by properties of the listener (e.g., Bent and Bradlow, 2003; Imai *et al.*, 2003; Labov and Ash, 1997; Mason, 1946) and linguistic context (e.g., Healy and Montgomery, 2007), as well as interactions among these factors (e.g., Moore, 2003; Rogers *et al.*, 2006).

Whether differences across talkers are maintained under different listening environments is an issue that has not been extensively studied. In one of the few extant studies, Cox *et al.* (1987) found that relative intelligibility rankings

[a)]Author to whom correspondence should be addressed. Electronic mail: tbent@indiana.edu

among six talkers were generally maintained across four levels of noise degradation (speech mixed with babble). More recently, Green *et al.*, (2007) reported no differences in talker intelligibility among three groups of listeners: normal-hearing listeners, CI listeners, and simulated CI listeners. The stimuli included single words appended to a carrier phrase from six talkers and semantically-anomalous sentences from two talkers. The stimuli were produced by two adult male, two adult female, and two child female talkers. Each of these three groups contained one high intelligibility talker and one low intelligibility talker (based on earlier results from Hazan and Markham, 2004). These stimulus materials were presented to CI users and normal-hearing listeners. Normal-hearing listeners heard the speech either mixed with multi-talker babble at a favorable signal-to-noise ratio or under CI simulation. Green *et al.* (2007) reported that intelligibility was relatively consistent across listeners and degradation types. These two studies suggest that at least some talker characteristics that promote intelligibility are beneficial across listener populations and listening conditions. The present study continues this line of inquiry by comparing the intelligibility of speech mixed with multi-talker babble to the intelligibility of CI-simulated speech for a larger number of talkers.

B. Perceptual learning

Several recent studies have found that listeners demonstrate both talker-dependent and talker-independent perceptual learning of speech (e.g., Bradlow and Bent, 2008; Norris *et al.*, 2003). With respect to talker-dependent learning, as listeners become more familiar with a talker's voice their word recognition accuracy increases (Bradlow and Bent, 2008; Nygaard *et al.*, 1994; Nygaard and Pisoni, 1998). These studies use overall intelligibility to assess adaptation, so it was not possible to determine how the specific experience with the talkers' voices enabled the listeners to improve their ability to identify those talkers' words. Other work using synthetic manipulations for specific phoneme contrasts suggests that listeners adjust their phonemic category boundaries in talker-specific ways (e.g., Eisner and McQueen, 2005; Norris *et al.*, 2003). These studies suggest that using lexical knowledge, listeners shift their category boundaries as needed for particular talkers.

The effect of linguistic experience has also been found to be talker-independent. Talker-independent learning has been shown for adjustments to phoneme category boundaries for native-accented speech (Kraljic and Samuel, 2006, 2007). Furthermore, a beneficial effect of experience on speech intelligibility has been shown for listeners with extensive experience listening to foreign accented speech (Bradlow and Bent, 2008; Clarke and Garrett, 2004; Weil, 2001), speech produced by talkers with hearing impairments (McGarr, 1983), speech synthesized by rule (Schwab *et al.*, 1985; Greenspan *et al.*, 1988), and computer manipulated speech (Dupoux and Green, 1997; Pallier *et al.*, 1998). Critically, this benefit has been reported to extend to new talkers and to new speech signals created using the same types of signal degradation (Bradlow and Bent, 2008; Francis *et al.*, 2007; Greenspan *et al.*, 1988; McGarr, 1983). Of particular relevance to the current study, listeners also show rapid adaptation to noise-vocoded speech (Davis *et al.*, 2005; Hervais-Adelman *et al.*, 2008). In the study of Davis *et al.* (2005) perceptual adaptation occurred without feedback across 30 sentences and was enhanced by feedback either in the form of an orthographic presentation of the stimulus or repetition of an unprocessed version of the stimulus. Adaptation was stronger with meaningful sentences than non-word sentences although adaptation with words was the same with real words and non-words (Hervais-Adelman *et al.*, 2008). Other training studies have also shown that the amount of perceptual learning depends on the type of materials listeners are trained with (Loebach and Pisoni, 2008).

Investigating listeners' adaptation to new talkers and the conditions that allow for this adaptation provides valuable information about the robustness and extent of plasticity in the speech perception system. Furthermore, uncovering the conditions that are most beneficial to adaptation can potentially help in the development of training programs for listeners with speech perception difficulties such as listeners with hearing impairment or second language learners. We addressed this type of perceptual adaptation in the present study by comparing performance on an initial group of sentences in a novel listening condition to performance after the listener has been exposed to the condition for many sentences. The results of previous studies suggest that we should observe significantly better performance after exposure to a novel listening condition. Furthermore, we investigated performance over the time-course of the experiment to determine the asymptote for adaptation by calculating performance in blocks of ten sentences (i.e., performance in the first ten sentences, second ten sentences, etc.).

While the studies reviewed above on perceptual learning of speech have demonstrated a great deal of flexibility of listeners' perceptual systems, they have typically focused on only one talker and only one type of signal degradation. In the current study, we addressed these gaps in two ways. First, we investigated adaptation across a large number of talkers to determine the extent of variation in perceptual adaptation to different talkers. Second, we compared how these differences in perceptual adaptation to different talkers may be affected by two different types of signal degradation. One type of signal degradation, CI-simulated speech, was selected because the perceptual adaptation of CI users is a topic that is still relatively unexplored, and attempts to understand their perceptual adaptation should be helpful in creating training protocol for individuals with CIs; synthesizing speech with a CI simulator for unimpaired subjects is a useful tool for addressing the issues with this population (Dorman and Loizou, 1998; Dorman *et al.*, 1997; Shannon *et al.*, 1995). The second type of signal degradation, mixing speech with multi-talker babble, was selected to be an ecologically valid degradation method that would provide a comparison with the CI-simulated speech; this will allow us to address whether individuals adapt to all types of signal degradation in the same way by determining whether the speakers to whom adaptation is more robust are the same in each condition. These forms of degradation are similar in that they both

make spectral detail less accessible. However, vocoding and the addition of multi-talker babble degrade the signal in different ways: spectral broadening and masking, respectively.

C. The present study

In this paper, we report on an investigation of how talker characteristics interact with degradation type to determine speech intelligibility and perceptual adaptation. One of the aims of this experiment was to determine whether and how inter-talker differences in intelligibility change depending on the type of degradation (i.e., CI-simulated speech versus speech mixed with multi-talker babble). The second aim of this study was to investigate how across talker differences and signal degradation type affect perceptual adaptation. Understanding how the interaction of talker characteristics and listening environment influences intelligibility and perceptual adaptation is an important goal in identifying the factors that contribute to speech perception and learning.

In the current experiment, intelligibility scores for ten male and ten female talkers were compared under two listening conditions: CI-simulation and multi-talker babble. Listeners were presented with speech from only one talker in one listening condition. Four hundred listeners were tested in total: 200 listeners for each listening condition. Intelligibility scores were compared across listening conditions, and the extent of adaptation to the speech over the time-course of the experiment was assessed.

II. METHOD

A. Stimuli

The experimental materials were sentences taken from the Indiana Multi-talker Sentence Database (Karl and Pisoni, 1994). This database includes recordings of 100 Harvard sentences (IEEE, 1969) produced by 20 talkers (10 male and 10 female), with a total of 2000 sentences. All talkers were speakers of general American English. The sentences were processed in two ways to assess speech intelligibility under CI-simulated listening conditions and when mixed with multi-talker babble.

B. CI-simulation

For the CI-simulation condition, each sentence was processed through an eight-channel sinewave vocoder using the CI simulator TIGERCIS (http://www.tigerspeech.com/). Stimulus processing involved two phases: an analysis phase, which used band pass filters to divide the signal into eight nonlinearly spaced channels (between 200 and 7000 Hz, 24 dB/octave slope) and a low pass filter to derive the amplitude envelope from each channel (400 Hz, 24 dB/octave slope), and a synthesis phase, which replaced the frequency content of each channel with a sinusoid that was modulated with its matched amplitude envelope. The eight-channel simulation was chosen because on average normal-hearing listeners perform similar to CI users when listening to eight-channel simulations compared to greater or fewer numbers of channels (Dorman *et al.*, 1997). Furthermore, a sine-wave vocoder was employed rather than noise-band vocoder because sine-wave vocoders also approximate CI user performance more closely than noise-band vocoders (Gonzalez and Oliver, 2005). However, it should be noted that this simulation is not an entirely accurate representation of the information presented to CI users. Specifically, due to the spectral side-bands around the sine-wave carriers, more information regarding the fundamental frequency is available in the simulation than is through a CI. The availability of this information may affect the intelligibility of speech in ways that are not representative of CI processing.

C. Multi-talker babble

For the multi-talker babble condition, the original sentences were mixed with six-talker babble (three male and three female talkers) at a signal-to-noise ratio of 0. The same babble file was used for each of the 2000 sentences. None of the talkers included in the babble file were the same as the target talkers. This signal-to-noise ratio was chosen based on pilot data in which the intelligibility of the sentences mixed with multi-talker babble was matched with intelligibility of the eight-channel CI-simulated sentences. The speech in this condition was not processed.

D. Participants

Four hundred normal-hearing listeners participated in this study (268 females and 132 males with an average age of 21.4 years). All listeners were native speakers of English and reported no current speech or hearing impairments at the time of testing. The majority of the participants were from the mid-west and indicated their place of birth as Indiana ($n=191$), Illinois ($n=50$), Ohio ($n=13$), Michigan ($n=13$), Minnesota ($n=6$), Missouri ($n=6$), Wisconsin ($n=3$), Iowa ($n=3$), or Kansas ($n=2$). The remaining participants were from the south ($n=40$), northeast ($n=29$), west ($n=21$), the U.S., state not specified ($n=6$), or outside of the U.S. ($n=7$). Ten participants did not indicate their place of birth. Most of the participants did not speak a foreign language, but 22 indicated knowing one language other than English. These languages included Spanish ($n=8$), Urdu ($n=3$), Chinese ($n=1$), French ($n=1$), German ($n=1$), Italian ($n=1$), Korean ($n=1$), Polish ($n=1$), Japanese ($n=1$), Swedish ($n=1$), and Arabic ($n=1$). Two of the participants indicated knowing two foreign languages: Hebrew/Spanish ($n=1$) and Bengali/Hindi ($n=1$). Listeners were either paid $5.00 for their participation or received course credit in an introductory psychology course. Participants were undergraduate students at Indiana University or members of the greater Bloomington community. In the CI-simulation condition, four subjects' data were removed because they were determined to be outliers (their keyword correct score was at least three standard deviations below the mean for that talker). Their data were replaced by data from four new listeners.

E. Experimental task

In each condition, a talker's intelligibility was assessed by examining the performance of ten normal-hearing listeners on a sentence transcription task (20 talkers $\times$ 2 degradation conditions $\times$ 10 listeners $=$ 400 listeners total). Each lis-

tener was presented with speech from one condition (i.e., CI-simulation or multi-talker babble) and heard only one talker during the course of the experiment, allowing us to assess differences in adaptation across talkers. During testing, each participant wore Beyer Dynamic DT-100 headphones while sitting in front of a Power Mac G4. Each sentence was played over the headphones followed by a dialog box presented on the screen, which prompted the listener to type what he or she heard. Each sentence was presented once in a randomized order. The experiment was self-paced so participants could take as long as needed to enter a response. Listeners were not provided with feedback as to the accuracy of their responses. Prior to the first experimental trial, participants were familiarized with the type of degradation by hearing two familiar nursery rhymes ("Jack and Jill" and "Star Light, Star Bright") produced by a talker not included in the Hoosier Multi-Talker Sentence Database or in the multi-talker babble, which had been processed in the same manner as the sentences in their experimental condition. During familiarization, listeners were not required to make any responses.

F. Scoring

The responses were scored based on number of keywords correct. Each test sentence has five keywords. Keywords were only counted as correct if all and only the correct morphemes were present. Words with added or deleted morphemes were counted as incorrect. Obvious misspellings and homophones were counted as correct.

III. RESULTS

The results under the two types of degradation were compared in several ways. First, intelligibility across talkers under the two types of degradation was compared in order to determine whether high and low intelligibility talkers in one condition are also the high and low intelligibility talkers in the other condition. Second, male speakers were directly compared with the female speakers in terms of intelligibility; gender was shown to be a significant predictor of intelligibility under quiet listening conditions (Bradlow *et al.*, 1996). Third, we examined the extent of perceptual adaptation under each experimental condition by comparing performance on the first 20 sentences with performance on the last 20 sentences. Differences in perceptual adaptation between the two conditions were also compared. We compared performance in ten blocks of ten sentences each to assess performance over the time-course of the experiment to investigate the rate of perceptual adaptation.

A. Comparison of intelligibility between the CI-simulation and multi-talker babble conditions

The intelligibility scores from the two conditions, CI-simulation and multi-talker babble, were compared. The keyword accuracy scores for the CI-simulated condition and the multi-talker babble condition were significantly correlated ($r=0.73$, $p<0.001$). Talkers who were highly intelligible under one type of degradation, CI-simulation, also tended to also be highly intelligible under the other type of degradation, multi-talker babble. A scatterplot of the keyword intelligibility scores in the two degradation conditions is shown in Fig. 1. It should be noted that different listeners were used for the two conditions. Therefore, potentially confounding listener variables were introduced (i.e., differences in dialect and other linguistic experiences across listeners).

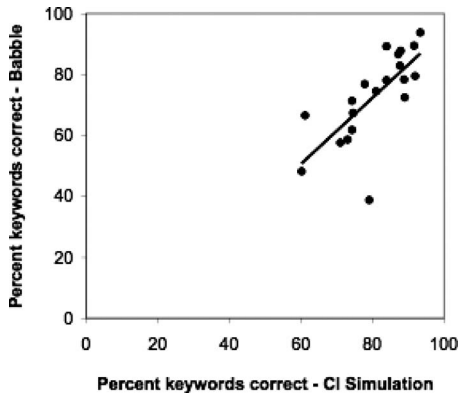


FIG. 1. Comparison of keyword intelligibility for the two degradation conditions, CI-simulated speech and speech mixed with multi-talker babble. Intelligibility scores under these two conditions were significantly correlated.

The intelligibility scores for each talker in the CI-simulation condition and the multi-talker babble conditions were also compared to intelligibility scores in the quiet (gathered by Karl and Pisoni, 1994). Here *sentence* intelligibility was considered rather than *keyword* intelligibility as Karl and Pisoni (1994) only reported sentence intelligibility scores (due to a lack of variation in keyword correct scores). Intelligibility scores in quiet were not significantly correlated with intelligibility in the CI-simulation condition ($r=0.35$, *ns*) and were not significantly correlated with intelligibility in multi-talker babble condition ($r=0.36$, *ns*). However, it should be noted that the range of intelligibility scores in the quiet was relatively small.

B. Gender differences

The data from both the CI-simulation and multi-talker babble conditions revealed that female talkers were more intelligible than male talkers. In the CI-simulation condition, female talkers (mean=84%, SD=11) were significantly more intelligible than male talkers [mean=77%, SD=11; $t(198)=4.61$, $p<0.001$]. Similarly, female talkers (mean=81%, SD=14) were more intelligible than male talkers [mean=65%, SD=13; $t(198)=8.47$, $p<0.001$] in the multi-talker babble condition. The gender difference in quiet, shown previously in Bradlow *et al.* (1996) with the same talkers, is maintained under the two forms of signal degradation tested here.

C. Perceptual adaptation

In addition to overall speech intelligibility, adaptation to the speech in each condition was assessed by examining improvement from the first 20 sentences to the last 20 sentences, a measure of perceptual adaptation. For the CI-simulation condition, this analysis revealed significant adaptation, with significantly more keywords correct in the

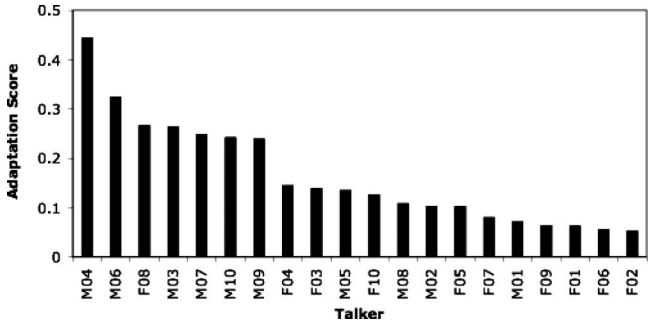


FIG. 2. Adaptation scores for the CI-simulated listening conditions. Talkers are ordered on the *x*-axis from left to right by their adaptation scores in the CI-simulation condition. While listeners adapted to the speech from all talkers, the extent of adaptation depended on the particular talker.

last 20 sentences (mean=84%, SD=11) than in the first 20 sentences [mean=73%, SD=15; $t(19)=16.6$, $p<0.001$]. Thus, listeners adapted to the CI-simulated speech without explicit feedback. An adaptation score was also calculated by subtracting the keywords correct in the first 20 sentences from the keywords correct in the last 20 sentences divided by keywords correct in the first 20 sentences. While listeners adapted to all talkers, a great deal of variation was observed in the extent of adaptation across individual talkers, with adaptation scores ranging from 0.05 to 0.44 for individual talkers. These data are shown in Fig. 2.

As with the CI-simulation condition, the perceptual adaptation analysis with the data from the multi-talker babble condition also revealed rapid adaptation, with significantly more keywords correct in the last 20 sentences (mean =75%, SD=13) than in the first 20 sentences [mean=69%, SD=16; $t(19)=6.45$, $p<0.001$]. Again, listeners rapidly adapted to the speech without explicit feedback. A great deal of variation was also observed in the extent of adaptation for the talkers, with adaptation scores ranging from 0.00 to 0.30 for individual talkers. These data are shown in Fig. 3.

In addition to assessing perceptual adaptation in each condition, we also compared the extent of adaptation in the two degradation conditions using the adaptation scores. A paired *t*-test revealed that listeners showed greater perceptual adaptation in the CI-simulated listening condition (mean =0.16, SD=0.11) than in the multi-talker babble condition

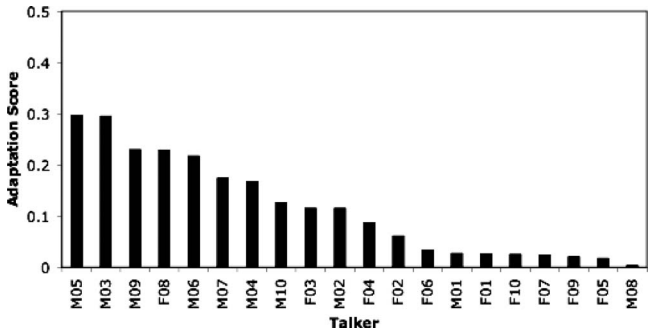


FIG. 3. Adaptation scores for speech mixed with multi-talker babble. The talkers are ordered on the *x*-axis based on their adaptation scores in the multi-talker babble condition, decreasing from left to right. While listeners adapted to the speech from all talkers, the extent of adaptation depended on the particular talker.

[mean=0.12, SD=0.10; $t(19)=2.68$, $p=0.015$]. Furthermore, when comparing each talker's adaptation scores between the two degradation conditions, the extent of adaptation in the two conditions was significantly correlated ($r=0.68$, $p=0.001$).

The vast majority of individual listeners showed improvement over the time-course of the experiment. In the CI-simulation condition, 91% of listeners showed adaptation scores above zero while in the multi-talker babble condition slightly fewer individuals showed improvement with 84% of adaptation scores above zero. Of the listeners who did not show adaptation across the course of the experiment (9% of listeners in the CI-simulation condition and 16% of listeners in the multi-talker babble condition), some listeners were at near ceiling level within the first 20 sentences, leaving little room for improvement over the time-course of the experiment. In the CI-simulation condition 7 of the 18 listeners who did not show perceptual adaptation were at 90% correct or above in the first 20 sentences and 12 of the 32 listeners showing no improvement in the multi-talker babble condition were at 90% or better in the first 20 sentences.

In addition to investigating performance during the beginning and the end of the experiment, we also examined the pattern of perceptual adaptation across the entire experiment by comparing performance for each block of ten sentences (i.e., Block 1=first ten sentences, Block 2=second ten sentences, etc.). This analysis allowed us to determine the point at which listeners reached asymptote in each condition and to compare the shape of the adaptation curves in each condition. The perceptual adaptation curves are shown in Fig. 4.

A repeated-measures analysis of variance (ANOVA) was conducted on these data with block as the within-subject repeated measure and condition (CI-simulation versus multi-talker babble) as the between-subjects variable. Results revealed main effects of block [$F(9,398)=77.84$, $p<0.001$] and condition [$F(1,398)=29.61$, $p<0.001$] as well as an interaction between block and condition [$F(9,398)=4.46$, $p<0.05$]. Because we found a significant interaction, separate repeated measures ANOVAs were conducted on the data from the two conditions.

For the CI-simulation condition, the effect of block was highly significant [$F(9,199)=60.96$, $p<0.001$]. *Post-hoc* pairwise comparisons with Bonferroni correction were made between each of the blocks in order to determine the asymptote. From these comparisons, listeners reached asymptote at the sixth block of sentences. That is, performance in the sixth block of sentences was not significantly different from any later blocks in the experiment, which also did not differ from one another. While listeners showed considerable adaptation across the first 60 sentences in the experiment (starting at 70% correct in the first ten sentences with gains to 83% correct in the sixth block of sentences), there was no further improvement observed after the sixth block (performance in the tenth block was only 1% higher than in the sixth block).

For the multi-talker babble condition, the effect of block was also highly significant [$F(9,199)=23.18$, $p<0.001$]. Pairwise comparisons revealed that listeners reached asymptote earlier in the experiment in this condition compared to the CI-simulation condition. In the multi-talker babble con-

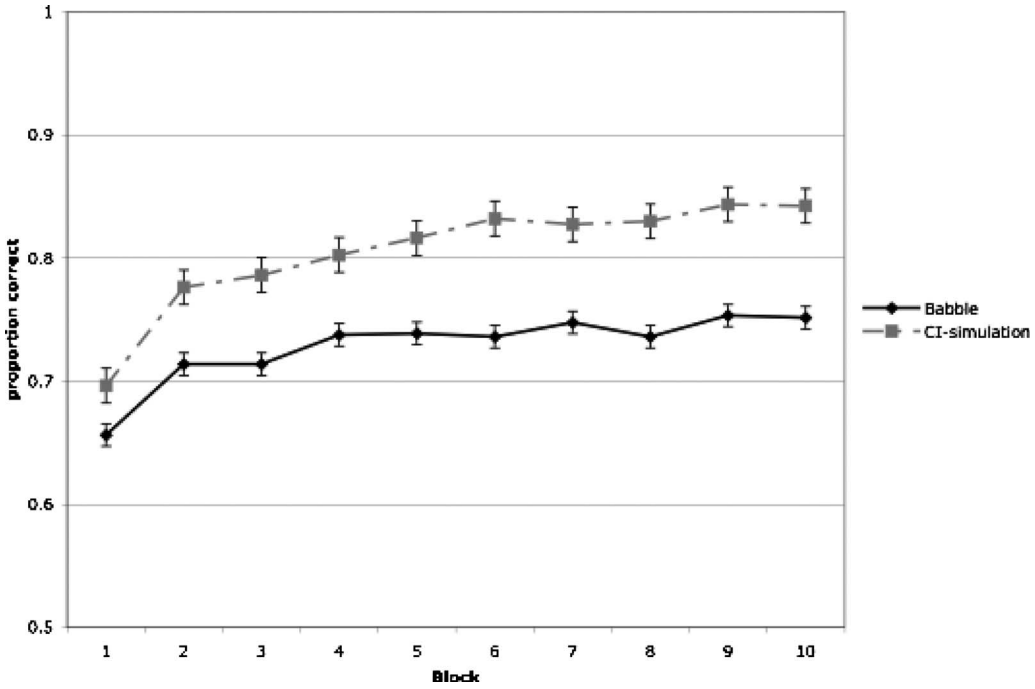


FIG. 4. Perceptual adaptation curves for the CI-simulation condition and multi-talker babble condition. On the *x*-axis performance for the ten blocks of sentences is shown (each composed of ten sentences). The *y*-axis displays proportion correct for keywords.

dition, Block 4 did not differ from any of the blocks later in the experiment, which also did not differ from one another. Listeners showed considerable adaptation in the first 40 sentences in the experiment (starting at 67% correct in the first ten sentences with gains to 74% in the fourth block of sentences). However, listeners showed little additional adaptation later in the experiment as performance only increased 1% from Block 4 (74%) to Block 10 (75%).

Comparisons were also made between the two conditions for each block using independent samples *t*-tests. Because of the large number of *t*-tests, Bonferroni correction was applied which indicated that *p*-values must be 0.005 or less to be considered significant. Comparisons across the two conditions showed that performance in the first block of ten sentences was not different in the two conditions [$t(398) = 2.12$, *ns*]. However, performance was significantly higher in the CI-simulation condition compared to the multi-talker babble condition in Blocks 2–10 ($p \leq 0.001$).

While the above analyses collapsed across male and female talkers, we wanted to investigate how learning across the experiment was affected by talker gender. Figure 5 shows learning across the time-course of the experiment in the two conditions divided by gender. It becomes clear that both initial and final performances are least accurate for male talkers when heard in multi-talker babble.

Results from the analyses of the perceptual adaptation revealed mostly similarities and some differences across the two degradation conditions. Adaptation scores across talkers were correlated for the two degradation conditions, and the adaptation effect was very robust. On average, listeners showed adaptation for nearly all talkers. Moreover, nearly all listeners showed adaptation over the time-course of the experiment. While the adaptation scores were correlated between the two conditions, listeners showed greater adaptation in the CI-simulation condition than in the multi-talker babble condition and showed the least accurate performance across the experiment for the male talkers in the multi-talker babble condition. Moreover, listeners reached asymptote later in the CI-simulation condition compared with the multi-talker babble condition.

IV. GENERAL DISCUSSION

Results from the current study suggest that across-talker differences in speech intelligibility are maintained across two types of signal degradation. Talkers who were found to be highly intelligible under CI-simulation were also highly intelligible when their speech was presented in multi-talker babble. Our findings support the recent conclusions of Green *et al.* (2007) who suggest that inter-talker differences are maintained across different listener groups (i.e., CI users, normal-hearing listeners presented with speech in a low level of babble or with CI-simulated speech). The present results replicated their earlier findings in a larger talker sample using sentence length materials. The overall patterns in our study diverge from previous studies examining relative intelligibility among talkers from different language backgrounds, indicating that speech intelligibility rankings may change depending on listener language background (Bent and Bradlow, 2003; Imai *et al.*, 2003; van Wijngaarden, 2001; van Wijngaarden *et al.*, 2002; cf. Major *et al.*, 2002; Munro *et al.*, 2006). However, we suspect that some factors such as language background may result in stronger talker-listener interactions compared with other factors such as hearing loss. If this is the case, then native listeners from the same speech community—regardless of their hearing status—will find the same talkers most intelligible, but listeners from different language backgrounds, especially na-

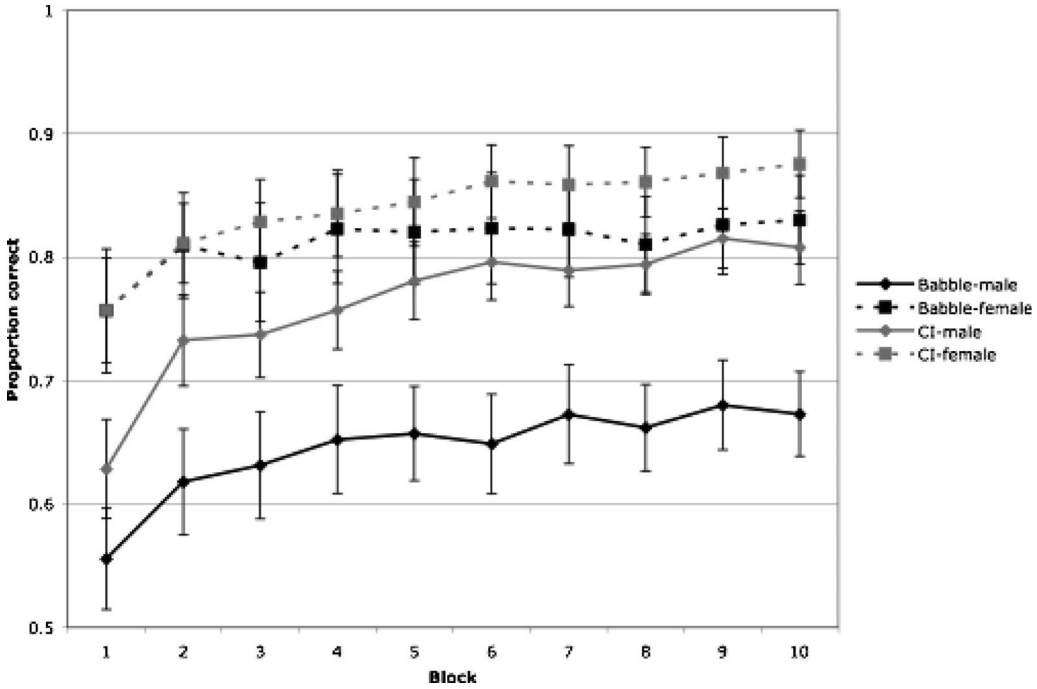


FIG. 5. Perceptual adaptation curves for the CI-simulation condition and multi-talker babble condition divided into male and female talkers. On the *x*-axis performance for the ten blocks of sentences is shown (each composed of ten sentences). The *y*-axis displays proportion correct for keywords.

tive and non-native listeners, may find different talkers most intelligible. The results from the present study reveal that intelligibility under multi-talker babble listening conditions is correlated with intelligibility under CI-simulation. However, as both of the degradation types tested here make the spectral detail in the speech signal less available, the extent to which this result can be generalized to other types of degradation remains an empirical issue. It should also be noted that different listeners were included in the two degradation conditions. As previous studies have found that factors about the listener can influence intelligibility, it may be the case that the results found here were influenced by across-listeners differences. A match in linguistic experiences of the talker and listener could have enhanced intelligibility scores for certain talkers, whereas a mismatch could have caused an intelligibility decrement. However, all listeners were native speakers of American English, and all talkers were speakers of general American English. Recent findings on the intelligibility of different American English dialects suggest that there is not an interaction between the dialect of the talker and the listener (Clopper and Bradlow, 2008).

Although mixing speech with multi-talker babble is typically considered an ecologically valid process for degrading speech, it should be noted that the same recordings—collected in quiet conditions—were used in the quiet and multi-talker babble listening conditions. Therefore, modifications that talkers make when they speak in noisy environments (e.g., Lombard speech: Junqua, 1993; Lane and Tranel, 1971; Lane *et al.*, 1970; Lombard, 1911; Summers *et al.*, 1988) were not performed in these recordings. In general, when listening in noise, speech produced in noise tends to be more intelligible than speech produced in the quiet (Summers *et al.*, 1988). Furthermore, certain talkers are more effective at making modifications and adjustments that help listeners in noisy environments when they are producing speech with noise present. Moreover, females generally tend to produce more intelligible Lombard speech than males (Junqua, 1993). Similarly, some talkers are better at making their speech highly intelligible when asked to speak clearly for listeners with hearing loss compared to the intelligibility of their speech when asked to speak conversationally (Ferguson, 2004; Picheny *et al.*, 1985; Uchanski *et al.*, 1996).

The remainder of this section explores two issues raised by the data reported here. In particular, we address the issues of perceptual adaptation and the observed gender differences.

A. Perceptual adaptation

Listeners with normal hearing are able to quickly adapt and accurately perceive speech under a variety of different listening conditions. In the present experiment, the analysis of adaptation to the degraded speech revealed the flexibility of the speech perception system. Even in the absence of any feedback, listeners recognized the talker's utterances more accurately after several minutes of exposure to the experimental stimuli (i.e., last 20 sentences) compared to the beginning of exposure to these stimuli (i.e., first 20 sentences). The extent of perceptual adaptation varied for each talker and in each type of signal degradation. It should be noted that for both degradation conditions, each listener in the experiment was only exposed to the speech of one talker. Therefore, the adaptation observed in the experiment is likely a result of adaptation to talker specific characteristics as well as adaptation to the degradation condition.

Listeners showed greater perceptual adaptation in the CI-simulation condition than in the multi-talker babble condition. However, the correlation between the talkers' adaptation scores in the two degradation conditions was positive

and significant. One likely source of the greater adaptation in the CI-simulation condition compared to the multi-talker babble condition is the novelty of the former type of degradation. The listeners in the current study had never experienced CI-simulated listening conditions before participating in the experiment, whereas listeners have had experience perceiving speech in environments with competing talkers. Thus, listeners are already practiced at picking out a given talker in noisy listening environments that are similar to the multi-talker babble condition, and must only adapt to the specifics of the multi-talker babble added to the speech in the experiment. Listeners are unlikely to learn a new listening strategy in this experiment, whereas the exposure to CI-simulated speech provided in this experiment was their first experience with this form of degradation. Therefore, they may have been able to acquire a new listening strategy during the course of the exposure to the CI-simulated speech. Evidence for this hypothesis comes from the finding that listeners continued to learn further into the experiment (i.e., they reached asymptote in the sixth block) in the CI-simulation condition compared with the multi-talker babble condition (i.e., they reached asymptote in the fourth block of sentences). The initial steep gains seen in both conditions may be a result of procedural learning while later learning may be a consequence of perceptual learning involving learning to better extract information from the degraded stimuli (Francis and Nusbaum, 2002). Another reason for the greater adaptation in the CI-simulation condition compared with the multi-talker babble condition is that the manipulation in the CI-simulation condition is a less variable and more predictable form of degradation than the multi-talker babble condition. Once a listener learns how the speech had been degraded in the CI-simulation condition, she can reliably use this information to more successfully interpret future utterances. In contrast, the way the multi-talker babble interacts with the target speech stimulus changes from sentence to sentence, which may hinder a listener's ability to apply knowledge learned from one sentence to the next. However, since the babble file that was mixed with the speech was the same from trial to trial, listeners may have been able to generalize their knowledge of the specifics of the babble noise from sentence to sentence (see Felty *et al.*, 2009). Although listeners showed robust learning for talkers in both conditions, the performance at the beginning and end of the experiment for the male talkers in the multi-talker babble condition was significantly lower than for female talkers in multi-talker babble or talkers from either gender in the CI-simulation condition.

It is worth noting that the listeners in this experiment did not receive any feedback, which suggests that they may have taken advantage of semantic and syntactic cues to enable them to learn how to perceive the speech under the two degradation conditions. Results from Davis *et al.* (2005) demonstrate that greater learning is observed in cases with meaningful sentences compared with non-word sentences, in which all words are non-words, or Jabberwocky sentences, in which only content words are replaced with non-words but real English function words remain. Davis *et al.* (2005) also tested perceptual learning of CI-simulated speech without feedback but only assessed 30 sentences. The present results add to their earlier findings by demonstrating that listeners continue to learn up through exposure to 60 sentences with eight-channel vocoded speech but then reach asymptote and show no further learning on the final 40 sentences. For listeners to achieve further gains, the inclusion of appropriate feedback would presumably be necessary.

In terms of generalizing about perceptual adaptation differences between babble and CI-simulation, it is worth noting that the experimental design employed here was a between-subjects design in which participants were exposed to only one type of signal degradation. While this methodological choice allowed us to fully explore the time-course and extent of perceptual adaptation to a given talker with a specific type of signal degradation as well as inter-talker differences, it does not allow us to definitively state whether one type of noise yields greater perceptual adaptation. Further research is required to explore this issue more fully.

The findings from the current study suggest that results from perceptual learning experiments using only one talker should be regarded with some caution, particularly with respect to the size of the perceptual learning effect. Most perceptual learning studies only use one talker or do not explicitly explore inter-talker differences. In line with our findings for degraded speech, Bradlow and Bent (2008) recently found that the extent of adaptation to foreign accented speech varied across talkers. Specifically, they found that the adaptation was greater for talkers with higher overall intelligibility. The issue of how perceptual learning is affected by overall intelligibility should be further explored with regard to the perception of degraded speech.

B. Talker gender

Previous studies have reported that adult female talkers are more intelligible than adult male talkers for normal-hearing adult and child listeners both in quiet and in low levels of noise (Bradlow *et al.*, 1996; Hazan and Markham, 2004) and for Lombard speech (Junqua, 1993). This result has been consistently observed across different types of materials (i.e., both words and sentences). The findings from the current study are consistent with these previous results and support the claim that female talkers tend to be more intelligible than male talkers in tests of talker intelligibility. The present study adds to the previous findings by demonstrating that this result holds under two types of signal degradation (e.g., under CI-simulation and with speech mixed with multi-talker babble). However, it should be noted that the same talkers were used in the current experiment as in the study of Bradlow *et al.* (1996)

The source of the gender difference is not known at this point. It is possible that female talkers are generally more intelligible than their male counterparts because of physical differences in the vocal tracts. However, the gender differences could stem from a learned source of behavior. For example, female talkers could make more extreme articulatory adjustments that result in more intelligible speech at the segmental level. If this latter type of explanation is the source of this difference, it would suggest that male talkers

could possibly be taught to alter their articulatory patterns to increase their intelligibility. Furthermore, it remains possible that women produce speech differently than men when being recorded by adopting a clearer speaking style even when not explicitly instructed to. More work is needed to resolve this issue.

V. CONCLUSIONS

The present results suggest that across-talker intelligibility differences are maintained under two types of signal degradation. High intelligibility talkers under CI-simulation also tended to be high intelligibility talkers in multi-talker babble listening conditions. These results replicate and extend the earlier intelligibility results of Green *et al.* (2007) by demonstrating that for a large number of talkers, intelligibility scores were significantly correlated for simulated CI listeners and normal-hearing listeners in noise. Furthermore, listeners were found to adapt rapidly to speech in both the CI-simulated and multi-talker babble conditions although greater perceptual adaptation was observed in the CI-simulation condition than in the multi-talker babble condition, and the extent of adaptation differed widely across talkers and listeners.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health to Indiana University (NIH-NIDCD T32 Grant No. DC-00012 and NIH-NIDCD R01 Grant No. DC-000111). An earlier version of this work was presented at the fourth Joint Meeting of Acoustical Society of America and the Acoustical Society of Japan, Honolulu, HI, November 2006. We thank Wesley Alford, Vidhi Sanghavi, Melissa Troyer, and Jennifer Karpicke for their assistance in data collection, Luis Hernandez for technical assistance, Larry Phillips for help with data entry, Ann Bradlow for allowing us access to her data, Jeremy Loebach for being so generous with his time, and Rochelle Newman and two anonymous reviewers for their many helpful suggestions.

[1]In this paper, we operationally define speech intelligibility as the listener's ability to accurately report the words that a talker has produced. This objective measure of speech intelligibility contrasts with other measures in which listeners subjectively rate the "intelligibility" of a speaker (also called comprehensibility; e.g., Fayer and Krasinski, 1987) or tests in which the listener must provide an accurate paraphrase of the talker's message for the talker's communicative intent to be considered effective (e.g., Brodkey, 1972).

Assman, P. F., and Summerfield, A. Q. (**2004**). "The perception of speech under adverse conditions," in *Speech Processing in the Auditory System*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. Fay (Springer-Verlag, New York).

Bent, T., and Bradlow, A. R. (**2003**). "The interlanguage speech intelligibility benefit," J. Acoust. Soc. Am. **114**, 1600–1610.

Black, J. W. (**1957**). "Multiple-choice intelligibility tests," J. Speech Hear Disord. **22**, 213–235.

Bond, Z. S., and Moore, T. J. (**1994**). "A note on the acoustic-phonetic characteristics of inadvertently clear speech," Speech Commun. **14**, 325–337.

Bradlow, A. R., and Bent, T. (**2008**). "Perceptual adaptation to non-native speech," Cognition **106**, 707–729.

Bradlow, A. R., Toretta, G. M., and Pisoni, D. B. (**1996**). "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics," Speech Commun. **20**, 255–272.

Brodkey, D. (**1972**). "Dictation as a measure of mutual intelligibility: A pilot study," Lang. Learn. **22**, 203–220.

Clarke, C. M., and Garrett, M. F. (**2004**). "Rapid adaptation to foreign-accented English," J. Acoust. Soc. Am. **116**, 3647–3658.

Clopper, C. G., and Bradlow, A. R. (**2008**). "Perception of dialect variation in noise: Intelligibility and classification," Lang Speech **51**, 175–198.

Cox, R. M., Alexander, G. C., and Gilmore, C. (**1987**). "Intelligibility of average talkers in typical listening environments," J. Acoust. Soc. Am. **81**, 1598–1608.

Davis, M. H., Johnsrude, I. S., Hervais-Ademan, A., Taylor, K., and McGettigan, C. (**2005**). "Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences," J. Exp. Psychol. Gen. **134**, 222–241.

Dorman, M., and Loizou, P. (**1998**). "The identification of consonants and vowels by cochlear implants patients using a 6-channel CIS processor and by normal hearing listeners using simulations of processors with two to nine channels," Ear Hear. **19**, 162–166.

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.

Dupoux, E., and Green, K. P. (**1997**). "Perceptual adjustment to highly compressed speech: Effects of talker and rate changes," J. Exp. Psychol. Hum. Percept. Perform. **23**, 914–927.

Eisner, F., and McQueen, J. M. (**2005**). "The specificity of perceptual learning in speech processing," Percept. Psychophys. **67**, 224–238.

Fayer, J. M., and Krasinski, E. (**1987**). "Native and non-native judgments of intelligibility and irritation," Lang. Learn. **37**, 313–326.

Felty, R., Buchwald, A., and Pisoni, D. B. (**2009**). "Adaptation to frozen babble in spoken word recognition," J. Acoust. Soc. Am. **125**(3), EL93–EL97.

Ferguson, S. H. (**2004**). "Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners," J. Acoust. Soc. Am. **116**, 2365–2373.

Fletcher, H., and Steinberg, J. C. (**1924**). "The dependence of the loudness of a complex sound upon the energy in the various frequency regions of the sound," Phys. Rev. **24**, 306–318.

Francis, A. L., and Nusbaum, H. C. (**2002**). "Selective attention and the acquisition of new phonetic categories," J. Exp. Psychol. Hum. Percept. Perform. **28**, 349–366.

Francis, A. L., Nusbaum, H. C., and Fenn, K. (**2007**). "Effects of training on the acoustic phonetic representation of synthetic speech," J. Speech Lang. Hear. Res. **50**, 1445–1465.

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.

Green, T., Katiri, S., Faulkner, A., and Rosen, S. (**2007**). "Talker intelligibility differences in cochlear implant listeners," J. Acoust. Soc. Am. **121**(6), EL223–EL229.

Greenspan, S. L., Nusbaum, H. C., and Pisoni, D. B. (**1988**). "Perceptual learning of synthetic speech produced by rule," J. Exp. Psychol. Learn. Mem. Cogn. **14**, 421–433.

Hazan, V., and Markham, D. (**2004**). "Acoustic-phonetic correlates of talker intelligibility in adults and children," J. Acoust. Soc. Am. **116**, 3108–3118.

Healy, E. W., and Montgomery, A. A. (**2007**). "The consistency of sentence intelligibility across three types of signal distortion," J. Speech Lang. Hear. Res. **50**, 270–282.

Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S., and Carlyon, R. P. (**2008**). "Perceptual learning of noise vocoded words: Effects of feedback and lexicality," J. Exp. Psychol. Hum. Percept. Perform. **34**, 460–474.

Hood, J. D., and Poole, J. P. (**1980**). "Influence of the speaker and other factors affecting speech intelligibility," Audiology **19**, 434–455.

Howes, D. (**1952**). "The intelligibility of spoken messages," J. Psychol. **65**, 460–465.

Howes, D. (**1957**). "On the relation between the intelligibility and frequency of occurrence of English words," J. Acoust. Soc. Am. **29**, 296–305.

IEEE (**1969**). "IEEE recommended practices for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 227–246.

Imai, S., Flege, J. E., and Walley, A. (**2003**). "Spoken word recognition of accented and unaccented speech: Lexical factors affecting native and non-native listeners," in *Proceedings of the International Congress on Phonetic Science*, Barcelona, Spain.

Junqua, J.-C. (**1993**). "The Lombard reflex and its role on human listeners

and automatic speech recognizers," J. Acoust. Soc. Am. **93**, 510–524.
Karl, J., and Pisoni, D. B. (**1994**). "The role of talker-specific information in memory for spoken sentence," J. Acoust. Soc. Am. **95**, 2873.
Kraljic, T., and Samuel, A. G. (**2006**). "How general is perceptual learning for speech?," Psychon. Bull. Rev. **13**, 262–268.
Kraljic, T., and Samuel, A. G. (**2007**). "Perceptual adjustments to multiple speakers," J. Mem. Lang. **56**, 1–15.
Labov, W., and Ash, S. (**1997**). "Understanding Birmingham," in *Language Variety in the South Revisited*, edited by C. Bernstein, T. Nunnally, and R. Sabino (University of Alabama Press, Tuscaloosa, AL).
Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709.
Lane, H., Tranel, B., and Sisson, C. (**1970**). "Regulation of voice communication by sensory dynamics," J. Acoust. Soc. Am. **47**, 618–624.
Loebach, J. L., and Pisoni, D. B. (**2008**). "Perceptual learning of spectrally degraded speech and environmental sounds," J. Acoust. Soc. Am. **123**, 1126–1139.
Lombard, E. (**1911**). "Le signe de l'elevation de la voix (The sign of elevating the voice)," Annales de Maladies d L'oreille et du Larynx **37**, 101–119.
Major, R., Fitzmaurice, S., Bunta, F., and Balasubramanian, C. (**2002**). "The effects of nonnative accents on listening comprehension: Implications for ESL assessment," TESOL Quarterly **36**, 173–190.
Mason, H. M. (**1946**). "Understandability of speech in noise as affected by region of origin of speaker and listener," Speech Monographs **13**, 54–68.
McGarr, N. S. (**1983**). "The intelligibility of deaf speech to experienced and inexperienced listeners," J. Speech Hear. Res. **26**, 451–458.
Miller, G. A. (**1947**). "The masking of speech," Psychol. Bull. **44**, 105–129.
Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perception confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.
Moore, B. C. J. (**2003**). "Speech processing for the hearing-impaired: Successes, failures and implication for speech mechanisms," Speech Commun. **41**, 81–91.
Munro, M., Derwing, T., and Morton, S. (**2006**). "The mutual intelligibility of foreign accents," Stud. Second Lang. Acquis. **28**, 111–131.
Norris, D., McQueen, J. M., and Cutler, A. (**2003**). "Perceptual learning in speech," Cogn. Psychol. **47**, 204–238.
Nygaard, L. C., and Pisoni, D. B. (**1998**). "Talker-specific learning in speech perception," Percept. Psychophys. **60**, 335–376.
Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (**1994**). "Speech perception as a talker-contingent process," Psychol. Sci. **5**, 42–46.
Pallier, C., Sebastian-Gallés, N., Dupoux, E., Christophe, A., and Mehler, J. (**1998**). "Perceptual adjustment to time-compressed speech: A cross-linguistic study," Mem. Cognit. **26**, 844–851.
Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1985**). "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech," J. Acoust. Soc. Am. **28**, 96–103.
Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., and Abrams, H. B. (**2006**). "Effects of bilingualism, noise and reverberation on speech perception by listeners with normal hearing," Appl. Psycholinguist. **27**, 465–485.
Schwab, E. C., Nusbaum, H. C., and Pisoni, D. B. (**1985**). "Some effects of training on the perception of synthetic speech," Hum. Factors **27**, 395–408.
Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (**1988**). "Effects of noise on speech production: Acoustic and perceptual analyses," J. Acoust. Soc. Am. **84**, 917–928.
Uchanski, R. M., Choi, S., Braida, L. D., Reed, C. M., and Durlach, N. I. (**1996**). "Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate," J. Speech Hear. Res. **39**, 494–509.
van Wijngaarden, S. J. (**2001**). "Intelligibility of native and non-native Dutch speech," Speech Commun. **35**, 103–113.
van Wijngaarden, S. J., Steeneken, H. J. M., and Houtgast, T. (**2002**). "Quantifying the intelligibility of speech in noise for non-native listeners," J. Acoust. Soc. Am. **111**, 1906–1916.
Weil, S. A. (**2001**). "Foreign-accented speech: Encoding and generalization," J. Acoust. Soc. Am. **109**, 2473.

On the assimilation-discrimination relationship in American English adults' French vowel learning[a)]

Erika S. Levy[b)]
Program in Speech and Language Pathology, Department of Biobehavioral Sciences, Teachers College, Columbia University, 525 West 120th Street, Box 180, New York, New York 10027

(Received 30 January 2009; revised 22 June 2009; accepted 15 August 2009)

A quantitative "cross-language assimilation overlap" method for testing predictions of the Perceptual Assimilation Model (PAM) was implemented to compare results of a discrimination experiment with the listeners' previously reported assimilation data. The experiment examined discrimination of Parisian French (PF) front rounded vowels /y/ and /œ/. Three groups of American English listeners differing in their French experience (no experience [NoExp], formal experience [ModExp], and extensive formal-plus-immersion experience [HiExp]) performed discrimination of PF /y-u/, /y-o/, /œ-o/, /œ-u/, /y-i/, /y-ɛ/, /œ-ɛ/, /œ-i/, /y-œ/, /u-i/, and /a-ɛ/. Vowels were in bilabial /rabVp/ and alveolar /radVt/ contexts. More errors were found for PF front vs back rounded vowel pairs (16%) than for PF front unrounded vs rounded pairs (2%). Overall, ModExp listeners did not perform more accurately (11% errors) than NoExp listeners (13% errors). Extensive immersion experience, however, was associated with fewer errors (3%) than formal experience alone, although discrimination of PF /y-u/ remained relatively poor (12% errors) for HiExp listeners. More errors occurred on pairs involving front vs back rounded vowels in alveolar context (20% errors) than in bilabial (11% errors). Significant correlations were revealed between listeners' assimilation overlap scores and their discrimination errors, suggesting that the PAM may be extended to second-language (L2) vowel learning. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3224715]

PACS number(s): 43.71.Hw, 43.71.An, 43.71.Es, 43.71.Gv [AJ] Pages: 2670–2682

I. INTRODUCTION

A predominant model of cross-language speech perception, the Perceptual Assimilation Model (PAM) (Best, 1995), posits that the perceived similarity of non-native segments to native categories, i.e., gestural constellations in native phonological space, predicts the difficulties naïve listeners will encounter in discriminating speech sounds in a non-native language. Exploring the extension of the PAM from the realm of naïve listeners to second-language (L2) learners, Best and Tyler (2007) proposed the PAM-L2 and called for research examining whether the principles involved in cross-language speech perception by naïve listeners also apply to L2 learning.

A limitation of the PAM (Best, 1995), the PAM-L2 (Best and Tyler, 2007), and of other speech perception and production models, such as Flege's (1995) Speech Learning Model (SLM), is that they are formulated qualitatively, with no objective measure of similarity between native and L2 speech sounds. The present study introduces the "cross-language assimilation overlap" method as a quantitative method for testing the claim by PAM and PAM L2 (Best, 1995; Best and Tyler, 2007) that perceived similarity of native and non-native (or L2) speech sounds predicts how accurately the non-native sounds will be discriminated. A study is reported on the discrimination of Parisian French (PF) front rounded vowels by native American English (AE) L2 learners of French. The discrimination results are compared to Levy's (2009) perceptual assimilation results by means of the cross-language assimilation overlap method.

In its original form, the PAM (Best, 1995) posits that naïve listeners perceptually assimilate speech sounds of an unfamiliar language into native categories and that their assimilation patterns predict the relative accuracy with which they will discriminate the segments. Non-native segments assimilate as gradiently "good" to "poor" instances of native categories along a continuum. In single-category assimilation, for example, segments that contrast in a non-native language are both assimilated as equally good or poor exemplars of the same native language category, yielding the highest degree of discrimination difficulty. In category-goodness assimilation, two non-native speech sounds are assimilated into the same native category, but one of the segments is perceived as a better instance than the other. In proposing the PAM-L2, Best and Tyler (2007) posited that when a category-goodness assimilation pattern occurs, there is little incentive for a new category to be learned for the less deviant L2 phone. The authors suggest that the deviant phone may be initially learned as a variant of the native category and that with continued L2 exposure, the language learner becomes more attuned to the relevant contrasts between the phones and creates a new L2 category. A factor in determining the creation of new L2 categories is whether the L2 contains minimally contrasting words that occur frequently in dense phonological neighborhoods, increasing the communicative necessity of perceiving the contrast.

[a)]Portions of this work were presented at the Acoustical Society of America meeting held in Providence, RI, in June, 2006.
[b)]Author to whom correspondence should be addressed. Electronic mail: elevy@tc.columbia.edu

In the PAM (Best, 1995) framework, two-category assimilation involves each non-native segment assimilating to a separate native category. An uncategorizable segment is assimilated within the native phonological space, but outside any native category. When the uncategorizable segment is paired with a segment that is similar to an AE category, an "uncategorized-categorized" assimilation pattern emerges. Both segments may also be uncategorizable in the native language. When two segments assimilate to separate native categories or as worse or better exemplars or as uncategorizable and categorizable exemplars, these patterns are expected to yield more accurate discrimination than pairs that fall into a single-category assimilation pattern. According to the PAM-L2 (Best and Tyler, 2007), if both L2 phones are assimilated in an uncategorizable pattern, learning depends, to some extent, on how similar the L2 phones are perceived to be to native phones that approximate them in phonological space.

Researchers have operationalized definitions of assimilation patterns referred to by the PAM (Best, 1995) in diverse ways. For example, Best *et al.* (2001) designated a non-native speech sound as "uncategorized" if a listener's orthographic transcription of the sound suggested one that fell between two or more native English categories. Other researchers (e.g., Levy, 2009; Strange *et al.*, 2009) have used inter- and intra-subject consistencies of categorization as indications of whether sounds are uncategorized. Harnsberger (2001) determined a speech sound to be uncategorized when its top label represented less than 90% of a group's responses. A limiting consequence of this classification method, which apportions continuous ranges into categories, is evident in those patterns referred to by Harnsberger (2001) as "borderline" cases. For example, if a group assimilates a non-native sound to a native category on 89% of trials and the other to another on 91% of another native category, this pattern is considered uncategorized-categorized, as it just misses criterion for the "two-category" or uncategorized-uncategorized patterns. Harnsberger (2001) responded to this type of problem by including borderline scores in more than one assimilation type (e.g., uncategorized-categorized and uncategorized-uncategorized) in his analysis.

Category goodness-of-fit ratings have also been relied on diversely in the field. For example, Best (1995) and Kuhl and Iverson (1995) found goodness ratings to be a strong predictor of discrimination accuracy. Guion *et al.* (2000) combined identification and goodness ratings into one metric in order to examine the relationship between cross-linguistic mapping patterns and discrimination. In contrast, Levy (2009), Strange *et al.* (2005, 2009) found that in phrase- and sentence-level non-native vowel perception experiments, listeners made use of a small range of goodness-of-fit ratings; thus, these studies made limited use of ratings in their analyses.

The various operational definitions of "categorization" and of "category goodness" may yield more than one way to classify assimilation patterns, thus leading to different predictions of discrimination accuracy. The cross-language assimilation overlap method introduced in this study was developed as a quantitative technique for examining perceptual assimilation and discrimination relationships. Rather than relying on the more typically used method of categorizing patterns according to type of perceptual assimilation (e.g., two-category, category-goodness, etc.) and then comparing expected performance based on perceptual assimilation type with actual performance (e.g., Best *et al.*, 1996, 1988; Harnsberger, 2001), this method ranks perceived similarity, which is quantified by an "overlap" score, and examines the correlation between listeners' overlap and their discrimination accuracy. Overlap is defined as the smaller percentage of responses when two members of a pair of non-native (or L2) speech sounds are assimilated to the same native category. This method permits the rank ordering of vowel contrasts in terms of difficulty predicted from perceptual assimilation patterns, without making reference to goodness ratings. A goal of the present study was to determine whether such an analysis would find a relationship between perceptual assimilation overlap and discrimination errors in French vowel learning.

It should be noted that the studies using perceptual assimilation and discrimination tasks in the PAM (Best, 1995) tradition have focused mostly on naïve listeners' performance (e.g., Best *et al.*, 1996, 1988; Best and Strange, 1992; Strange *et al.*, 2001). Few experiments thus far (e.g., Guion *et al.*'s [2000] study of consonant perception) have examined L2 learners' discrimination patterns. To the author's knowledge, none has been used to examine vowel perception by experienced learners, even though accurate vowel recognition has been found to be more important than consonant recognition for overall sentence intelligibility (Kewley-Port *et al.*, 2007).

II. THE DISCRIMINATION EXPERIMENT

This section reports a study of the effects of formal and immersion language experience and consonantal context on AE listeners' discrimination of PF contrasts involving front rounded vowels. French high front rounded /y/ and mid front rounded /œ/[1] are produced with the tongue forward and the lips protruded (Tranel, 1987). English, in contrast, has no canonical front rounded vowels, although in several AE dialects, /u/, /ʊ/, and /o/ have become more "fronted," i.e., produced with the tongue farther forward in the oral cavity (Clopper *et al.*, 2005; Strange *et al.*, 2007). Findings are mixed regarding AE speakers' discrimination of front rounded vowels from other French vowels. High accuracy is reported in Best *et al.*'s (1996) categorial[2] discrimination study, in which naïve AE listeners discriminated Bretagne French /sœ-sy/ syllables with fewer than 5% errors. Similarly, in a study involving L2 learners, Flege and Hillenbrand (1984) tested native English speakers proficient in French on paired /tu-ty/ tokens produced by seven native French speakers from France and Belgium. Listeners identified which member of the pair was /ty/ with an error rate of only 10%.

Greater problems in discrimination of front rounded vowels were found for even advanced AE learners of French in Gottfried's (1984) categorial discrimination study. AE listeners with and without French experience and native French listeners heard productions of PF vowels /e-ɛ/, /a-ɛ/, /i-ɛ/,

/ɑ-ɔ/, /y-u/, /a-ɑ/, /y-ø/, and /œ-ø/ in /tVt/, /Vt/, /tV/, and /V/ syllabic contexts, uttered as if in sentences. Vowels in isolation were discriminated more accurately than vowels in /tVt/ context by all three groups.

In the reviewed studies, the vowel stimuli were presented either in alveolar context or in isolation. Production studies indicate that vowels vary depending on their consonantal contexts (Hillenbrand *et al.*, 2001) and that patterns of variation differ in different languages (Strange *et al.*, 2007), suggesting that learning coarticulatory patterns of variation may be part of the L2 speech learning process (Beddor *et al.*, 2002; Levy and Law II, 2008; Manuel, 1999; Oh, 2008). Phonetic context may affect vowel perception (Bohn and Steinlen, 2003), as well. Strange *et al.* (2009) found effects of consonantal context and speaking style (i.e., citation form disyllables vs sentences) on assimilation of French and German vowels in sentences by naïve AE listeners. For example, PF /y/ was more often assimilated to AE /u/ in alveolar (94%) than in bilabial (74%) context.

In an investigation of context effects in L2 learning, Levy and Strange (2008) extended Gottfried's (1984) study, examining AE listeners' discrimination of PF vowels /y/, /œ/, /u/, and /i/ in /rabVp/ and /radVt/ bisyllables in AXB triads of the phrase "neuf /raCVC/ à des amis," ("nine /raCVC/ to some friends"). (In the AXB paradigm, stimuli are presented in triads, with the second matching the first or the third.) Two groups of AE listeners participated: The "inexperienced group" consisted of AE listeners with no French experience. The "experienced group" was highly proficient in French, with extensive classroom and immersion French experience. Results showed effects of French language experience and consonantal context on AE listeners' discrimination of the French contrasts. The experienced group made fewer errors (5%) than the inexperienced group (24%) for three experimental pairs (PF /y-i/, /œ-u/, and /y-œ/). For PF /y-u/, no statistically significant difference (24% for Inexp vs 30% for Exp) was revealed as a function of language experience, pointing to this contrast as a particularly difficult one to master. The inexperienced group made more PF /y-u/ errors in alveolar context (8% in bilabial vs 39% in alveolar), but more PF /y-i/ errors in bilabial context (24% in bilabial vs 8% in alveolar). The experienced group confused PF /y-u/ in both contexts (24% in bilabial and 35% in alveolar context) with great between-subject variation. For all contrasts except PF /y-i/, where the opposite was the case, the inexperienced group made fewer errors in bilabial than in alveolar context. No significant context effect was found for the experienced group.

An explanation for the context-dependent performance by these L2 learners makes reference to the relationship between native and L2 vowel production. High back AE vowels /u/, /ʊ/, and /o/, to a lesser extent, are fronted in alveolar contexts (Hillenbrand *et al.*, 2001; Strange *et al.*, 2007). (See Fig. 1 in Levy, 2009, for a vowel plot of the PF vowel stimuli superimposed onto AE vowel space.) Thus, in AE, the phonological category /u/ has relatively back rounded [u] instantiations in most contexts (e.g., "cool" [kul]), but when the tongue is forward, as in alveolar context in "dude" [dyd], for example, AE /u/ approximates a front rounded vowel. Thus,

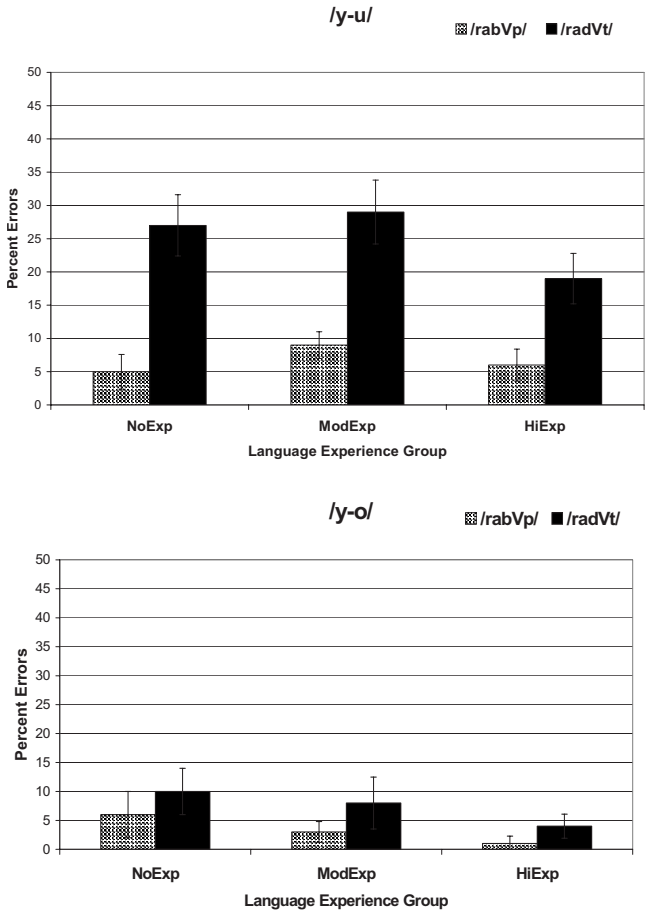


FIG. 1. Language and context effects on /y/ discrimination. Categorial discrimination of PF /y-u/ (top) and /y-o/ (bottom) in bilabial (/rabVp/) and alveolar (/radVt/) contexts by AE listeners with no French experience (NoExp), moderate French experience (ModExp), and extensive French experience (HiExp): percent errors and standard errors.

to native speakers of languages with front rounded vowel categories, [u] and [y] represent two different phonological categories. In English, on the other hand, the segments [y] and [u] may be allophones of the phonological category /u/. AE listeners, then, may tend to perceive high and mid front rounded vowels as more similar to AE /u/ or /ʊ/ (in which fronting can be expected) when the segments are in alveolar context than when they are in other contexts. Thus, they may confuse front rounded vowels with back PF vowels that also assimilate to back AE categories, especially in alveolar context.

The patterns with which L2 learners assimilate vowels as a function of L2 experience and consonantal context were investigated in a perceptual assimilation study by Levy (2009). AE listeners with no French experience (NoExp group), AE listeners with formal French classroom learning experience, but no immersion (ModExp), and AE learners with extensive classroom and immersion experience (HiExp group) participated. Listeners performed an assimilation task involving PF /y, œ, u, o, i, ɛ, a/ in bilabial /rabVp/ and alveolar /radVt/ contexts, presented in phrases. They were given a choice of 13 AE key words ("heed, hid, hayed, head, had, hod, hawed, hud, hoed, hood, who'd, hued, and herd") and were asked to select the word that contained the vowel

most similar to the target. They rated the vowel on a scale from 1–9 ("most foreign-sounding" to "most English-sounding").

Levy (2009) found that front rounded vowels were assimilated primarily to back AE vowels (PF /y/ to palatalized AE /ʲu/, and PF /œ/ to AE /ʊ/). Back rounded PF vowels were also assimilated to back AE vowel categories (PF /u/ to AE /u/, and PF /o/ to AE /u/ and /o/). No language experience effect was found for PF /y/ to AE /ʲu/ assimilation in alveolar context (NoExp=65%, ModExp=71%, and HiExp =61%). In bilabial context, on the other hand, listeners with extensive experience assimilated PF /y/ to AE /ʲu/ less often (72%) than listeners with no (80%) or only formal (85%) experience. For PF /œ/, assimilation patterns differed as a function of language experience and consonantal context. With extensive experience, /œ/ assimilated more often to AE /ʊ/ (e.g., in alveolar context, NoExp=17% and HiExp =61%) or /ɝ/ (e.g., in alveolar context, NoExp=0% and HiExp=20%) and less to /u/ (e.g., in alveolar context, NoExp=59% and HiExp=9%). Both front rounded vowels assimilated more often to AE /u/ in alveolar context (e.g., for PF /y/ assimilation to AE /u/, NoExp=31%) than in bilabial context (NoExp=7%).

The PAM (Best, 1995) predicts poor discrimination of contrasts that assimilate in a single-category pattern. Hence, according to the PAM, Levy's (2009) finding of perceptual assimilation of front rounded vowels to back vowels was consistent with AE listeners' greater difficulty distinguishing front rounded vowels from back rounded vowels than from front unrounded vowels reported by Levy and Strange (2008). However, AE listeners' assimilation of front rounded vowels to back vowels was not consistent with Best's *et al.*'s (1996) and Flege and Hillenbrand's (1984) finding of relatively high accuracy in /y-u/ discrimination in naïve listeners and L2 learners. These inconsistencies may be thought of in terms of the Automatic Selective Perception model of speech perception (Strange and Shafer, 2008), which posits that L1 selective perceptual routines are relied on to a greater extent when task demands increase. It is possible that the tasks in Levy and Strange's (2008) discrimination study and Levy's (2009) assimilation study, involving vowels in phrases uttered by three speakers in continuous speech, were more demanding than tasks in earlier studies using citation materials uttered by a single speaker, for example, yielding poorer perceptual outcomes.

The present study investigated the effects of French language experience and consonantal context on AE listeners' discrimination of L2 French vowels, extending Levy and Strange's (2008) discrimination study in three ways: First, an additional group of listeners (ModExp) with just classroom experience was tested. Second, for a more comprehensive examination, discrimination of additional vowel pairs (front rounded vs back rounded /y-o/ and /œ-o/; and front rounded vs front unrounded /y-ɛ/, /œ-ɛ/, and /œ-i/) was targeted as well as the four also examined by Levy and Strange (2008), i.e., front rounded vs front unrounded /y-i/, front rounded vs back rounded, /y-u/ and /œ-u/, and front rounded vs front rounded /y-œ/. And finally, the same participants whose assimilation data were collected for Levy (2009)[3] were tested on discrimination in order for assimilation-discrimination relationships to be examined, as described in Sec. III.

This study investigated (1) whether AE L2 learners of French had more difficulty discriminating PF front rounded vowels from PF front unrounded vowels or from PF back rounded vowels, (2) whether level of French experience affected the listeners' discrimination accuracy, and (3) whether consonantal context affected their discrimination accuracy.

Because previous studies indicate that the discrimination of front vs back rounded vowels tends to be more difficult for AE listeners than does the discrimination of front rounded PF vowels vs other front vowels (Gottfried, 1984; Levy and Strange, 2008), more front vs back rounded vowel confusions were expected in the present experiment than front rounded vs unrounded vowel confusions. Overall, it was predicted that experienced L2 learners would demonstrate more accurate discrimination of contrasts than would inexperienced learners (Levy, 2009; Levy and Strange, 2008). Specifically, the HiExp group was expected to perform more accurately on the categorial discrimination task than the ModExp group, who was expected to perform more accurately than the NoExp group. However, based on the findings of Gottfried (1984) and Levy and Strange (2008), even the most-experienced AE listeners were predicted to have difficulty with the /y-u/ contrast. Other pairs expected to be difficult for the less-experienced listeners were /y-œ/ and /œ-o/, i.e., front rounded vowels paired with each other and front rounded vowels paired with vowels of a similar height. Consonantal context was expected to have a significant effect on discrimination, especially for inexperienced listeners, with front rounded vowels being less accurately discriminated in alveolar than in bilabial context.

A. Method

1. Stimulus materials

The stimulus materials for this study were identical to those described by Levy (2009). In brief, three female adult native PF speakers who had lived in the United States for less than a year were recorded as they read nine PF vowels, blocked by bilabial /rabVp/ or alveolar /radVt/ context in the sentence: "J'ai dit neuf /raCVC/ à des amis." (I said nine /raCVC/ to some friends.) A Shure microphone fed the signal to a Soundblaster Live Wave sound card via an Earthworks microphone preamp. The digital files were segmented so that only "neuf /rabVp/ à des amis" and "neuf /radVt/ à des amis" remained, with the target front rounded /y, œ/ and /i, u, ɛ, o, a/ for comparison. Task verification was accomplished by three monolingual native PF speakers visiting the United States for less than a month, who performed the categorial discrimination task (described below). They made no (0) errors on the experimental pairs, a total of three errors (=3% errors per pair) on the non-experimental pairs PF /u-i/, /y-ɛ/, and /y-o/ and reported that they had no difficulty performing the task.

An acoustic analysis of the PF stimuli was conducted by Levy (2009) and compared to AE acoustic values in Strange *et al.*'s (2007) production study. Although a full description is beyond the scope of this article, the following should be

noted: In bilabial context, PF /y/ approximated PF and AE /i/ far more than it approximated PF and AE /u/. PF /œ/ was intermediate between front AE vowels and back AE vowels. In alveolar context, PF /y/ still approximated PF /i/ more than it approximated PF /u/. However, in alveolar context, both PF and AE vowels /u/ and /o/ were fronted compared to their counterparts in bilabial context. PF /œ/ was only slightly fronted in this context. Thus, if the naïve and experienced participants had more difficulty discriminating PF /y/ from back than from front vowels, acoustics alone could not explain their patterns.

2. Participants

The three groups of participants in this experiment were the same as those described by Levy (2009). All 39 participants were raised in monolingual English-speaking households in the United States. The NoExp group was comprised of 13 native AE speakers, ages 20–40 years, who were living in New York City and had never studied French, nor lived in a French-speaking country, nor interacted significantly with French speakers. The ModExp group were 13 native AE speakers, ages 22–37 years, who were living in New York City, and had studied French in classroom settings, but had minimal French immersion experience. They had started learning French at a mean age of 16.1 years (SD=2.8) for 2–4 years (mean=3 years and SD=0.8). They had not lived in a French-speaking country for more than 5 months. The HiExp group were 13 native AE speakers, ages 20–61 years, with extensive classroom and immersion French experience, who were speaking French regularly (range=2 h/week –100% of the time, median=15 h/week). They had studied French for a mean of 8 years (range=5–13 years and SD =2.4), starting no earlier than age 12 years (mean age of starting=14 years and SD=1.6). They had spent at least 1 year living in a French-speaking country in adulthood (range=1–16 years and median=1.4 years), and spoke French frequently around the time of the experiment. Participants passed a hearing screening at 20 dB.

3. Procedure

Participants listened to the discrimination stimuli presented by STAX Professional SR Lambda headphones connected to an amplifier (STAX Professional SRM-1/MK-2), receiving the signal from the Dell Dimension XPS B800 computer in a sound-attenuated chamber. The five experimental "one-feature" vowel pairs presented were PF /y-i/, /y-u/, /œ-ɛ/, /œ-o/, and /y-œ/. These were contrasts whose members differed in just one feature (e.g., rounded vs unrounded for PF /y-i/ or front vs back for PF /y-u/, high vs mid for PF /y-œ/). The six "two-feature" vowel pairs were PF /y-ɛ/, /œ-i/, /y-o/, /œ-u/, /u-i/, and /a-ɛ/, whose members differed by more than one feature (e.g., back rounded vs front unrounded for PF /u-i/). In addition, all of these pairs included at least one vowel with a "counterpart"[4] in AE (/i, u, ɛ, o/). The two-feature pairs PF /u-i/ and PF /a-ɛ/ were considered control pairs because they had counterparts in AE and did not include front rounded vowels. Two-feature vowel pairs were expected to be more accurately discriminated overall by virtue of being phonologically "more different" than the one-feature pairs in PF. Four orders were possible for presentation of each A-B vowel pair: AAB, ABB, BBA, and BAA. Trials contained triads of stimuli uttered by three different speakers in random order, blocked by consonantal context, with an equal number of correct A and B responses. Conditions were counterbalanced such that all nine tokens of each vowel occurred in each contrasting pair an equal number of times.

The stimuli were randomized and presented using the "Paradigm Discrim" program (by Bruno Tagliaferri). The pairs were arranged into AXB trials. Subjects were instructed to click on "1" if the vowel in the second stimulus was the same vowel in the first, and "3" if it was the same as the vowel as in the third. Prior to testing, AE subjects were given task familiarization in which they were asked to discriminate 18 trials of vowel pairs involving AE /ɛ/, /ɑ/, /œ/, and /ɪ/ vowel pairs in the AXB paradigm. Participants were permitted no more than two errors on task familiarization in order to continue with the experiment. All participants met these criteria.

The AE task familiarization was followed by French stimulus familiarization. Stimulus familiarization was identical to the experimental task. Following the stimulus familiarization in one context, listeners heard 4 blocks of 24 experimental trials in that context, then 1 block of stimulus familiarization trials in the other context, followed by 4 blocks of 24 experimental trials in that context. Each listener completed 12 judgments for each of the five one-feature pairs in each consonantal context, resulting in 60 one-feature trials per context. Six judgments were completed for each of the six two-feature pairs, resulting in 36 judgments on the two-feature pairs. Thus the experiment consisted of a total of 96 triads in each context. The inter-stimulus interval was 500 ms and trials were self-paced.

B. Results

1. Data analysis

Discrimination scores were derived by tallying errors over trials for each contrast in each context and converted to percentages of errors of total number of trials. An error was defined as responding 3 when the trial was AAB or 1 when the trial was ABB.

2. Language experience and consonantal context effects

For an overview of categorial discrimination findings, Table I presents the percent errors by each language experience group (NoExp, ModExp, and HiExp across the top row) for each contrasting vowel pair, with consonantal contexts combined. The discrimination scores for the vowel pairs are listed beginning with scores for the experimental pairs, followed by the control pair scores. The individual experimental pairs are discussed in Secs. II B 3–II B 6 with regard to the language experience and consonantal context effects revealed. The overall discrimination errors for the experimental pairs decreased with language experience (mean=13%, 11%, and 3% for NoExp, ModExp, and HiExp, respectively).

TABLE I. Categorial discrimination of PF vowel pairs summed over /rabVp/ and /radVt/ contexts by AE listeners with no French experience (NoExp), moderate French experience (ModExp), and extensive French experience (HiExp): Percent errors and standard error of the mean (in percent) are given.

PF vowel pairs (Expt.)		No Exp % Error	Mod Exp % Error	Hi Exp % Error
High front rounded vs back rounded	y-u	16	19	12
	y-o*	8	5	3
Mid front rounded vs back rounded	œ-o	33	22	4
	œ-u*	33	29	6
High front rounded vs front unrounded	y-i	7	1	0
	y-ɛ*	1	1	1
Mid front rounded vs front unrounded	œ-ɛ	3	2	1
	œ-i*	0	3	1
High front rounded vs mid front rounded	y-œ	17	16	4
Control pair	a-ɛ*	19	10	2
Control pair	u-i*	1	1	0

*Note that vowel pairs with asterisks were two-feature vowel pairs and were presented for 12 judgments per participant. (The others were one-feature pairs, presented for 24 judgments.)

Vowel pairs were not equally difficult to discriminate, with the NoExp group making 0%–33% errors, depending on which contrast was presented.

For an analysis of whether PF front rounded vowels were more often confused with PF back rounded or PF front unrounded vowels, the discrimination data were divided into two scores: Total percent of errors made on pairs containing a front rounded vowel and a front unrounded vowel (PF /y-i/, /y-ɛ/, /œ-ɛ/, and /œ-i/) and total percent errors made on pairs containing a front rounded vowel and a back rounded vowel (PF /y-u/, /y-o/, /œ-o/, and /œ-u). When front rounded and unrounded vowels were contrasted, listeners in all groups made few errors (3%, 2%, and 1% for the NoExp, ModExp, and HiExp groups, respectively). When front and back rounded vowels were contrasted, on the other hand, listeners made far more errors (22%, 19%, and 6% for the NoExp, ModExp, and HiExp groups, respectively), as predicted.

Because listeners in all three groups made almost no errors on front rounded vs unrounded pairs, the remaining analyses focused on discrimination of front rounded vowels paired with back rounded vowels and with each other. On the four front rounded vs back rounded pairs (PF /y-u/, /y-o/, /œ-o/, and /œ-u/), the NoExp group made the most errors (22%), followed by the Mod Exp group (19%), followed by the HiExp group (6%). Because of heterogeneity of variance, nonparametric statistics were performed. As described below, a Kruskal–Wallis one-way analysis of variance (ANOVA) was implemented to examine the language experience effects on each of the four vowel pairs and Mann–Whitney U-tests provided pairwise comparisons for language experience and consonantal context.

3. Discrimination of PF /y-u/ and /y-o/: Language experience and consonantal context

Figure 1 presents mean errors for discrimination of pairs involving the front rounded vowel /y/ contrasted with the two back rounded vowels /u/ and /o/. The top graph shows percent errors (*Y*-axis) in discrimination of the /y-u/ pair by the NoExp, ModExp, and HiExp Groups (along the *X*-axis). Scores for each language group are divided into bilabial (left checkered bar) and alveolar (right solid bar) contexts. For the high vowel pair PF /y-u/, listeners performed above chance, but not significantly differently across groups: NoExp =16%, ModExp=19%, and HiExp=12%. A Kruskal–Wallis one-way ANOVA by language group confirmed that the language experience effect was not statistically significant [p =0.22]. This is consistent with Levy and Strange's (2008) finding that advanced listeners of French fared no better than listeners with no French experience for this vowel pair—a contrast that is particularly resistant to improvement. A Mann–Whitney U-test revealed the consonantal context main effect to be statistically significant [U=252, $p<0.001$] at a two-tailed significance level, on the other hand, as predicted from Levy's (2009) perceptual assimilation findings, with more difficulty revealed in alveolar context than in bilabial context.

The bottom graph in Fig. 1 presents the data for the PF /y-o/ contrast. As the figure shows, few errors were made by any group on this pair (NoExp=8%, ModExp=5%, and HiExp=3%). With so few errors, no significant experience effect [p=0.21] or consonantal context effect [U=656, p =0.15] was present.

4. Discrimination of PF /œ-o/ and /œ-u/: Language experience and consonantal context

Figure 2 presents discrimination results for pairs involving /œ/ and back rounded vowels. For the PF /œ-o/ contrast (upper graph), the NoExp group made the most errors (26% in bilabial and 39% in alveolar context), followed by the ModExp group (15% in bilabial and 29% in alveolar context), followed by very few errors by the HiExp (1% in bilabial and 6% in alveolar context). A Kruskal–Wallis one-way ANOVA by language group revealed a main effect of language experience [$p<0.001$], with increased experience being associated with fewer errors in discrimination for this vowel pair. A Mann–Whitney U-test indicated that the ModExp group made significantly fewer errors than did the NoExp group (U=45, p=0.04, two-tailed); thus, formal in-

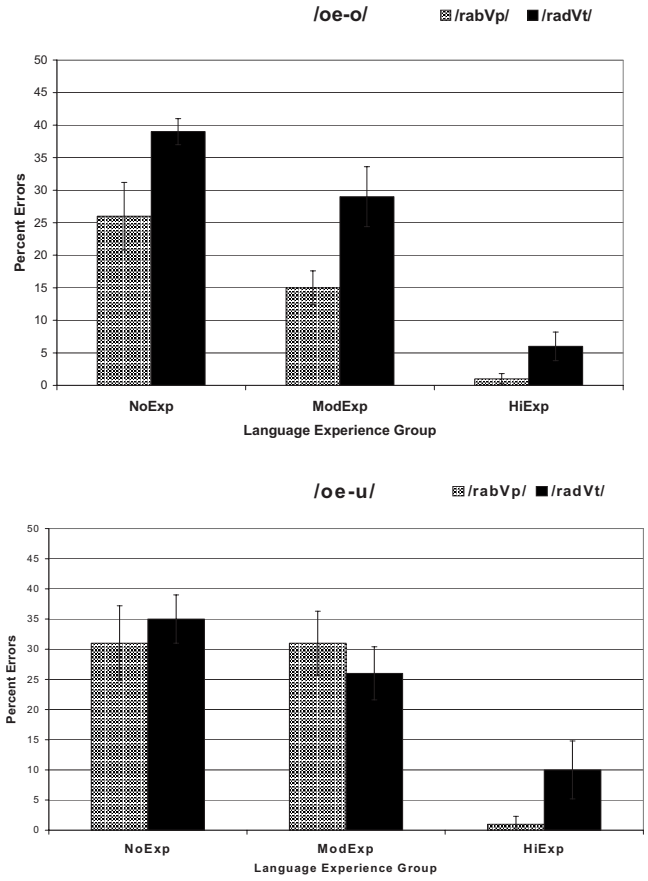


FIG. 2. Language and context effects on /œ/ discrimination. Categorial discrimination of PF /œ-o/ (top) and /œ-u/ (bottom) in bilabial (/rabVp/) and alveolar (/radVt/) contexts by AE listeners with no French experience (NoExp), moderate French experience (ModExp), and extensive French experience (HiExp): percent errors and standard errors.

struction was associated with (marginally) more accurate discrimination for this vowel pair. The HiExp group performed significantly more accurately than did the ModExp group ($U=10$, $p<0.001$); thus, extensive language instruction and immersion were associated with fewer errors than was formal instruction without immersion. The prediction of a consonantal context effect, based on assimilation differences as a function of context for PF /œ/, was also borne out ($U=505$, $p<0.01$, two-tailed), with more errors in alveolar context than in bilabial for all groups.

The bottom graph in Fig. 2 displays discrimination results for the PF /œ-u/ contrast. A Kruskal–Wallis one-way ANOVA by language group indicated a main effect of language experience [$p<0.001$]. The NoExp group made the most errors (31% in bilabial and 35% in alveolar context), followed by the ModExp group (31% in bilabial and 26% in alveolar context), followed by very few errors by the HiExp (1% in bilabial and 10% in alveolar context). Thus, as predicted, a language effect was present, with increased experience associated with fewer errors in discrimination of this vowel pair. However, for this pair only, the immersion group performed more accurately than the other groups ($U=13$, $p<0.001$, two-tailed). The formal experience group (ModExp) did not perform significantly more accurately than the NoExp group ($U=13$, $p=0.39$, two-tailed). An unexpected finding for this pair was the lack of a significant context effect [$U=688$, $p=0.45$], despite differences in assimilation of /œ/ as a function of context.

5. Interaction of vowel pair, language group, and consonantal context for PF /y-u/ and /y-i/

When consonantal context was taken into consideration, the only score to reach above 6% errors in pairs involving front rounded vs unrounded vowels was the score of 10% errors for the PF /y-i/ pair in bilabial context by the NoExp group. Despite the low error rate, this contrast merits examination in light of the interaction of vowel pair, language experience, and consonantal context.

In Levy (2009), a subgroup of NoExp listeners perceptually assimilated PF /y/ to AE front unrounded /i/. (The other language experience groups rarely assimilated /y/ to /i/ in either context.) An interaction was found in the present study, primarily for one individual with no French experience, in which PF /y/ was assimilated to AE /i/ more often in bilabial context than in alveolar context. The interaction in discrimination is consistent with the interaction found in assimilation in Levy (2009). In bilabial context, the NoExp group made more errors for PF /y-i/ (10%) than for PF /y-u/ (5%), whereas in alveolar context, they made more errors for the PF /y-u/ pair (27%) than for the PF /y-i/ pair (4%). As in the assimilation task, this pattern was primarily due to one participant, who made 33.3% errors on PF /y-i/ in bilabial and 0% errors in alveolar context.

A closer examination of that listener's perceptual assimilation and discrimination patterns provides an example of the PAM (Best, 1995) or the PAM-L2 (Best and Tyler, 2007) being predictive on an individual level: The NoExp listener perceptually assimilated all PF /y/ vowel stimuli in bilabial context to AE /i/ (100% of responses—more than all other listeners). As predicted by the PAM, he had discrimination difficulty (33% errors) with the PF /y-i/ contrast in bilabial context—the highest percentage of errors of any participant on this pair. In alveolar context, on the other hand, he perceptually assimilated PF /y/ exclusively to back vowels (39% to AE /u/, 50% to AE /ʊ/, and 6% to AE /ʲu/—never to AE /i/). As predicted, in alveolar context, he discriminated PF /y-i/ far more accurately (0% errors) than the PF /y-u/ contrast (25% errors). Thus, for this individual listener, the PAM predicted discrimination performance from perceptual assimilation patterns, and most effectively when consonantal context was taken into account.

6. Discrimination of PF front rounded vowels /y-œ/: Language experience and consonantal context

As shown in Fig. 3, NoExp listeners made the most errors in differentiating the PF /y-œ/ pair (12% errors in bilabial and 22% in alveolar context), followed by ModExp (10% errors in bilabial and 21% in alveolar context), followed by HiExp (3% errors in bilabial and 4% in alveolar context). A Kruskal–Wallis one-way ANOVA by language group indicated a main effect of language experience [$p<0.001$]. A Mann–Whitney U-test indicated no significant difference between performance of the NoExp group and the

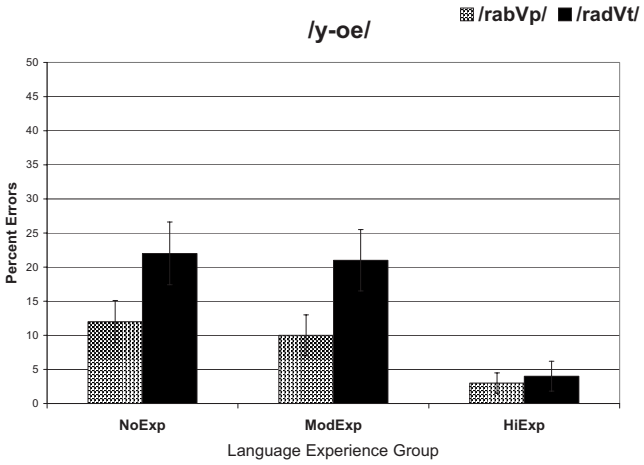


FIG. 3. Categorial discrimination of PF front rounded vowels /y-œ/ in bilabial (/rabVp/) and alveolar (/radVt/) contexts by AE listeners with no French experience (NoExp), moderate French experience (ModExp), and extensive French experience (HiExp): percent errors and standard errors.

ModExp group, [$U=77$, $p=0.69$], but a significant difference between ModExp and HiExp performance, [$U=21$, $p<0.001$]. According to a Mann–Whitney U-test, consonantal context only approached statistical significance [$U=572$, $p=0.05$] for this vowel pair.

7. Discrimination of control pairs PF /a-ɛ/ and PF /u-i/

The control pairs PF /a-ɛ/ and PF /u-i/ were, by definition, expected to result in few discrimination errors, based on the assumption that these vowels would fall into a two-category assimilation pattern. For the PF /a-ɛ/ pair, the groups made more errors than expected (NoExp=19%, ModExp=10%, and HiExp=2%). Levy (2009) indicated that the NoExp group perceived both PF /a/ and /ɛ/ as most similar to AE /œ/ some of the time; thus, it appears that listeners without immersion experience may have assimilated these segments in a single-category assimilation pattern instead. The control contrast PF /u-i/ was indeed discriminated without difficulty by all language groups (NoExp=1%, ModExp=1%, and HiExp=0% errors), indicating that listeners were on task.

C. Discussion

In summary, AE listeners had more difficulty discriminating PF front rounded vowels from PF back rounded vowels than from PF front unrounded vowels. Overall, listeners who had formal-plus-immersion experience with French performed significantly more accurately than those without L2 French experience and those with only formal French instruction experience. Only the PF /y-u/ vowel pair remained relatively difficult for highly experienced listeners. Discrimination of pairs involving the mid front rounded vowel /œ/ with back rounded vowels was more accurate with greater L2 experience, especially with extensive formal-plus-immersion experience. Listeners made more errors with the /œ-u/ pair than with the /œ-o/ pair, despite the height difference in the first pair. For the /y-œ/ pair, an experience effect was also evident in the non-immersion vs immersion groups.

Overall, discrimination of front vs back rounded vowels and discrimination of front rounded vowels from each other was significantly less accurate in alveolar context than in bilabial context. The context effect was evident in both pairs involving vowels of a similar height (PF /y-u/ and /œ-o/). These findings had been predicted based on previous literature, including Levy's (2009) assimilation results. Contrary to expectations, no context effect was found for front rounded vowels paired with vowels of a different height. For the PF /y-o/ pair, this may be attributed to too few errors to reveal a significant interaction. The lack of a context effect in the PF /œ-u/ pair is less interpretable.

Stimuli in nearly all previous studies of the perception of front rounded vowels by AE listeners have been vowels preceded and/or followed by alveolar consonants (e.g., Best *et al.*, 1996; Flege, 1987; Flege and Hillenbrand, 1984; Gottfried, 1984; Polka, 1995; Polka and Bohn, 1996) or produced in isolation (e.g., Gottfried, 1984; Rochet, 1995; Stevens *et al.*, 1969). Results from the present experiment suggest that replications of such studies, but using stimuli in which the vowels are produced in other consonantal contexts, may reveal different results. In bilabial context, for example, AE listeners are likely to make fewer discrimination errors for pairs involving front rounded vowels than indicated in previous literature, although some naïve individuals may have more difficulty discriminating the PF /y-i/ contrast in bilabial than in alveolar context.

The overall effect of more accurate discrimination by the HiExp group than by the ModExp was not true for every vowel pair. Both formal and immersion experience in late L2 learners were associated with increased accuracy in perception of non-native contrasts. For the PF /y-u/ vowel pair, formal experience alone was not associated with greater accuracy, consistent with Levy's (2009) finding of no experience effect for perceptual assimilation of PF /y/ to AE /ʲu/ in alveolar context (a pattern that would predict two-category assimilation, thus higher discrimination accuracy), but not with the finding of an experience effect of decreased PF /y/ to AE /ʲu / assimilation by the HiExp group in bilabial context. In the present study, listeners immersed in French for several years performed essentially the same as those with no French experience, lending support to studies that point to the PF /y-u/ contrast as one particularly resistant to perceptual learning by AE listeners (e.g., Gottfried, 1984; Levy and Strange, 2008).

Compared to naïve listeners, listeners with formal training alone discriminated the PF mid-vowel pair /œ-o/ more accurately, and listeners with extensive formal training and immersion performed with the greatest accuracy. No higher discrimination accuracy for the PF vowel pairs /œ-u/ and /y-œ/ was associated with merely formal training, but extensive training and immersion were associated with significantly more accurate discrimination. That discrimination accuracy with formal instruction only was not greater than with no French exposure supports the notion that, to be most effective, language instruction programs must include more than the typically administered foreign language requirements in United States schools.

III. TESTING THE PAM ON L2-VOWEL LEARNING

In Sec. II, discrimination results were, for the most part, predicted based on the perceptual assimilation patterns reported in Levy (2009), with confusions arising when the front rounded and back rounded PF vowels in a pair assimilated the same back AE categories, which occurred most often in alveolar context. On an individual level, it was shown that a participant with no French experience, who assimilated PF /y/ to front vowels in bilabial context and to back vowels in alveolar, also had more difficulty discriminating PF /y-i/ in bilabial context than in alveolar context.

This section reports a more systematic examination of discrimination accuracy for the vowel pairs tested in relation to the same listeners' assimilation patterns, accomplished through the cross-language assimilation overlap method. This method was used to examine the relationship (i.e., correlation) between degree of overlap in assimilation (i.e., how often two non-native vowels perceptually assimilated to the same native category) and the discriminability of vowel pairs in order to test the predictions generated by the PAM (Best, 1995) for L2 vowel learning (Best and Tyler, 2007).

That is, by quantifying perceptual assimilation overlap (e.g., for the PF /y-u/ pair, how often tokens of both PF /u/ and PF /y/ assimilated to the same AE vowel category /u/), it was possible to place contrasting pairs along a continuum from most similar to least similar. This permitted more finely grained predictions to be made about relative discrimination accurately for vowel pairs. Additionally, for the purposes of the present study, it was not evident how to characterize the assimilation of AE /u/ and /^{j}u/ response categories in Levy (2009). It was not clear whether this was two-category (palatalized /u/ vs nonpalatalized /u/), category goodness (allophonic variation), or single-category (phonological /u/) perceptual assimilation.

It was hypothesized that (1) vowel pairs whose members assimilated to separate categories (by each language experience group) would be discriminated more accurately (by the same language experience group) than those pairs whose members assimilated to the same categories, an outcome predicted by the PAM (Best, 1995) for naïve listeners and the PAM-L2 (Best and Tyler, 2007) for L2 learners, and that (2) the more trials in which an individual assimilated both members of a vowel pair to the same native category (i.e., the higher the overlap score), the less accurate the individual's discrimination would be for that vowel pair. Both hypotheses were expected to be true for all three language experience groups and in both consonantal contexts.

A. Cross-language assimilation overlap by language experience group

1. Data analysis

In testing the first hypothesis, that vowel pairs whose members assimilated to separate categories would be more discriminable than those whose members did not, the cross-language assimilation overlap method proceeded as follows:[5] Vowel pairs were the sampling variable. For this analysis, an overlap score was obtained for each vowel pair within each language group. The overlap was operationally defined as the smaller percentage of responses when two members of a PF pair were perceptually assimilated to a particular AE vowel category. For the /y-u/ experimental vowel pair in bilabial context, for example, when PF /u/ was presented to NoExp listeners in the perceptual assimilation task, the modal response (90.2%) was /u/. When /y/ was presented, 6.8% of stimuli were categorized as /u/; thus for 6.8% of the stimuli (the portion that overlaps between 90.2% and 6.8%, i.e., the smaller percentage), perception of /y/ and /u/ overlapped. In addition, the NoExp group categorized PF /u/ as /^{j}u/ for 6.4% of the stimuli, which overlapped with the modal choice of /^{j}u/ (79.9%) when PF /y/ was presented. Both /y/ and /u/ also were perceived as closest to /i/ for an overlap of 0.4%, and both were perceived as /ʊ/ for an overlap of 1.7%. Thus, when 6.8%, 6.4%, 4%, and 1.7% (the overlap percentages when both stimuli were assimilated to the same AE vowel) were summed, the result was a total overlap score of 15.3% for the perception of /y-u/ by the NoExp group in bilabial context.

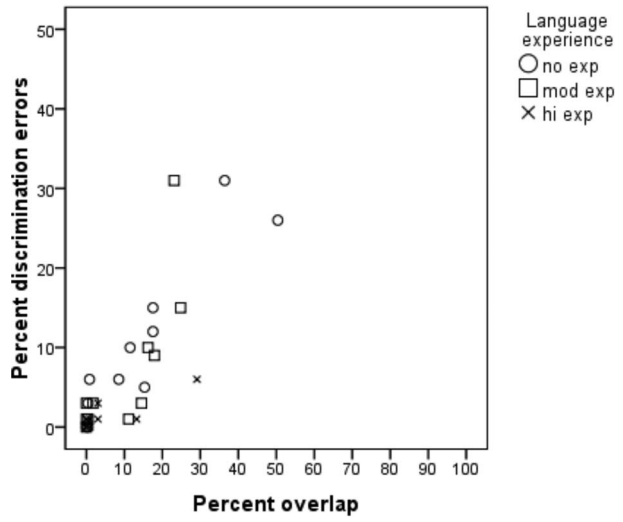
Overlap scores were tallied for the remaining experimental vowel pairs in each consonantal context within each language group and then ranked from lowest to highest. Finally, the discrimination error scores associated with each vowel pair were correlated with total overlap score for each pair. (In the above example, the NoExp group made 5% discrimination errors for /y-u/ in bilabial context, which was compared with the percentage overlap score of 15.3% for that pair.) Nonparametric correlations (Spearman rank order) were performed because the perceptual assimilation results could not be considered interval measures. The higher the overlap score, the higher the percent errors were expected to be revealed in discrimination. Thus, when overlap scores for each pair were ranked from lowest to highest, discrimination error results were also predicted to be ordered from lowest to highest.

2. Results

The Appendix lists the cross-language assimilation overlap score and the categorial discrimination percent errors for each language experience group, arranged by overlap score (in ascending order) for each group in each consonantal context. A pattern of more discrimination errors with higher overlap scores is evident, with contrasts involving front rounded vowels paired with back vowels and with each other generally revealing more overlap and more discrimination errors than the other pairs. For example, for NoExp listeners in alveolar context, the scores ranged from 65.9% overlap and 39% discrimination errors (for PF /œ-o/) to 0% overlap and 0% discrimination errors (for PF /œ-i/).

Figure 4 graphs the correlation between cross-language assimilation overlap and discrimination performance for vowel pairs in bilabial /rabVp/ context (A) and in alveolar /radVt/ context (B). Along the *x*-axis are the cross-language assimilation overlap scores, while the *y*-axis represents the percent errors in discrimination (up to chance of 50%). Each point on the graph represents a group's response to a particular vowel pair. Data for the NoExp Group are represented as Xs, whereas data for the ModExp group are represented by squares, and for the HiExp group, by circles. As shown in

A. Bilabial context



B. Alveolar context

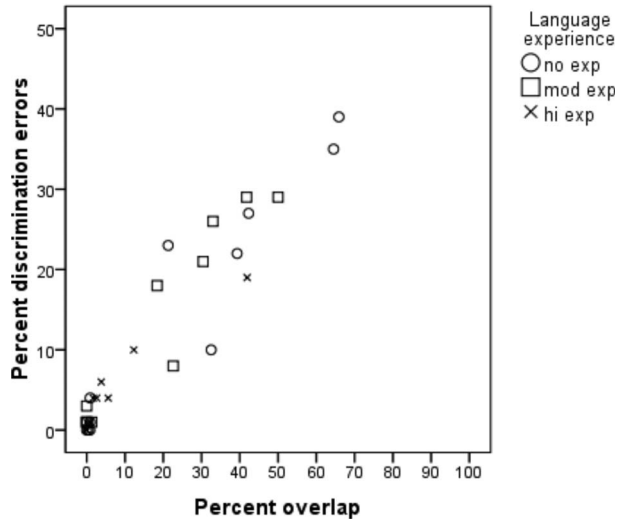


FIG. 4. Scatterplot of relationship between cross-language assimilation overlap patterns and percent errors in categorial discrimination in bilabial /rabVp/ context (a) and in alveolar /radVt/ context (b) by AE listeners with no French experience (no exp), moderate French experience (mod exp), and extensive French experience (hi exp) with vowel pairs /y-u/, /œ-o/, /y-o/, /œ-u/, /y-i/, /y-ɛ/, /œ-ɛ/, /œ-i/, /a-ɛ/, /u-i/, and /y-œ/ as sampling variables.

Fig. 4(A), for the NoExp group, as perceptual overlap increased in bilabial context, so did discrimination errors. A Spearman rank order correlation confirmed a strong correlation between overlap scores and discrimination errors ($\rho = 0.92$, $p < 0.001$). Thus, for this naïve group of listeners, perceptual assimilation patterns were highly predictive of discrimination performance on French vowel contrasts in bilabial context, as posited by the PAM (Best, 1995).

For the ModExp group, the correlation for bilabial context data was also statistically significant ($\rho = 0.84$, $p = 0.001$). Thus, for these L2 learners with formal French instruction, the PAM (Best, 1995) and PAM-L2 (Best and Tyler, 2007) successfully predicted relative vowel discrimination difficulty. Results were also significant for the HiExp group ($\rho = 0.68$, $p < 0.05$). However, this correlation is not particularly informative, as there were so few errors to interpret for that group. As the figure shows, most of the Xs representing the HiExp group cluster and overlap below 6% in perceptual assimilation and below 4% errors in discrimination, with only one outlying vowel pair, PF /y-u/, revealing higher overlap (29%) and discrimination error (6%) scores.

Turning to alveolar context, Fig. 4(B) reveals more variability in discrimination errors and perceptual assimilation overlap (i.e., larger spread) for all groups than was seen in bilabial context, reflecting the perceptual difficulties encountered by AE listeners in this context. With more errors to work with, the correlation coefficients were higher than for the bilabial context comparison for the three language experience groups (NoExp $\rho = 0.96$, $p < 0.001$; ModExp, $\rho = 0.93$, $p < 0.001$; and HiExp $\rho = 0.95$, $p < 0.001$). Thus, in support of the first hypothesis, the PAM (Best, 1995) and PAM-L2 (Best and Tyler, 2007) predicted relative accuracy in discrimination of vowel pairs from their assimilation patterns, not only for naïve learners of a language, but also for intermediate and advanced adult L2 learners.

B. Cross-language assimilation overlap by individuals

1. Data analysis

A further correlational analysis was conducted in order to test hypothesis 2, that individuals' difficulty in discriminating a vowel pair would be related to their cross-language assimilation overlap of that vowel pair's members. As opposed to the previous analysis in which vowel pairs were ranked according to their group overlap scores, in this analysis, individuals' overlap scores in each consonantal context for PF /y-u/, /œ-u/, /y-œ/, and /œ-o/ were ranked and compared to their discrimination scores for each of these contrasts. These vowel pairs were chosen as these were the contrasts involving front rounded vowels that continued to pose the most difficulties discrimination (i.e., ≥10% discrimination errors in at least one consonantal context by ModExp) despite language experience, presenting an opportunity to test the PAM (Best, 1995) quantitatively for individual L2 learners (Best and Tyler, 2007).

For this analysis, each listener's perceptual overlap score was tallied for each vowel contrast, defined here as the percent of times that, given a particular vowel contrast, the listener perceptually assimilated both vowel pair members to the same native phone. For example, on the /œ-o/ contrast in alveolar context, a ModExp listener perceptually assimilated PF /œ/ to AE /o, ʊ, u, ʌ/ categories on 22%, 39%, 28%, and 11%, of trials, respectively. This listener assimilated PF /o/ to AE /o, ʊ, u/ on 67%, 17%, and 17% of trials, respectively. To calculate the overlap score for this listener for this vowel pair in alveolar context, the smaller percentages when both PF speech sounds assimilated to the same AE category (/o/ =22%, /ʊ/=17%, /u/=17%, and /ʌ/=0%) were summed (=56%). The overlap score was compared to that listener's discrimination score for the same vowel pair in the same consonantal context (in this case 42% errors).

2. Results

A Spearman rank order correlation confirmed a correlation between individuals' overlap scores and discrimination errors for PF /y-u/, /œ-u/, /y-œ/, and /œ-o/ ($\rho=0.68$, $p<0.001$, two-tailed) with all groups and both consonantal contexts included. Correlations were statistically significant for each language experience group when consonantal contexts were combined (NoExp $\rho=0.67$, $p<0.001$; ModExp $\rho=0.46$, $p<0.001$; and HiExp $\rho=0.58$, $p<0.001$) and for each consonantal context when language experience groups were combined (bilabial $\rho=0.60$, $p<0.001$; alveolar $\rho=0.69$, $p<0.001$). Results indicated that the more often (i.e., the more trials in which) individuals assimilated two members of a PF vowel pair to a single native category, the more discrimination errors they incurred for that vowel pair, providing quantitative support for the PAM (Best, 1995) and its extension to L2 learners (Best and Tyler, 2007).

C. Discussion

1. Quantifying perceptual assimilation patterns

The cross-language assimilation overlap method provided a measure of the accuracy of PAM (Best, 1995) and PAM-L2 (Best and Tyler, 2007) predictions. This method examined the frequency with which two members of a vowel pair both assimilated to a particular native category and compared that frequency to the same group's or individual's discrimination accuracy for the pair in question. This method revealed that, generally, the more often L2 vowels in a pair were assimilated to the same native category by a particular group or individual, the less accurately the contrast was discriminated by that group or individual, results that support the PAM's position that discriminability is predictable from assimilation patterns.

This method does not necessarily differentiate categorizable-uncategorizable patterns from two-category (or from uncategorizable-uncategorizable) assimilation patterns in that it examines only whether one vowel in a pair assimilated to the same native category as did the other vowel. Moreover, it does not capture within-category differences. For example, it does not factor in the goodness ratings that differentiate single-category from category-goodness assimilation patterns. However, as listeners used a limited range of ratings in Levy (2009), and as the assimilation types were not self-evident in this study, the absence of ratings information was not expected to affect the findings meaningfully and the lack of reliance on assimilation types may have been an advantage of the quantification method for this study.

2. Limitations

A limitation of the study is that, although cross-speaker tasks were implemented, stimuli were uttered by only three native PF speakers; thus, the overlap and discrimination scores obtained from the listeners' responses may not be generalizable. However, as the predictions were tested based on assimilation and discrimination responses to the same data set, replications of this study are expected to yield different scores, but similar relationships between cross-language assimilation overlap and discrimination accuracy.

It should also be noted that response choice alternatives in an assimilation task may affect response patterns (in this quantification and in more traditional qualitative methods) and, thus, the resulting correlations with discrimination results. For example, had AE /^{j}u/ not been a response alternative in Levy (2009), listeners might have assimilated more PF /y/ stimuli to AE /u/ than they did with the palatalized option, resulting in greater overlap in for the PF /y-u/ pair.

And finally, despite the significant correlations found in the present study, categorization models may not capture the complexity of non-native listeners' perceptual sensitivity, as demonstrated by Iverson *et al.*'s (2008) study of the categorization of English /w-v/ by native speakers of Sinhala, German, and Dutch speakers. Native speakers of Sinhala and German have one native phoneme similar to English /w/ and /v/, yet German speakers discerned the English /w-v/ distinction more successfully. Listeners were clearly sensitive to distinctions that were not necessarily reflected in their categorization patterns. Iverson *et al.* (2008) suggested that distortions in perceptual space also contribute to L2 learning.

IV. GENERAL DISCUSSION

Examining AE listeners' overlap in perceptual assimilation of non-native and L2 PF vowels in relation to the same listeners' discrimination errors yielded significant correlations in bilabial and alveolar contexts. These findings provide preliminary support for predictions generated by the PAM (Best, 1995) to be extended to the domain of listeners in the more advanced stages of L2 learning, as proposed by Best and Tyler (2007). As this was the first study to test the PAM's predictions on vowel learning using a quantitative measure, replication of such studies using the cross-language assimilation overlap and other methods is needed to support this conclusion.

Findings from the reported discrimination experiment and other studies (e.g., Gottfried, 1984; Levy, 2009; Levy and Strange, 2008; Strange *et al.*, 2001, 2009) suggest that contextual variation in the phonetic realization of vowels across languages impacts L2 vowel learning. Levy's (2009) assimilation study indicates that PF /y/ will be perceived as most similar to AE /u/ more often in alveolar context than in bilabial context, leading to more PF /y-u/ discrimination difficulty in alveolar context (with more assimilation overlap) than in bilabial context, as found in the present study. Similarly, when surrounded by bilabials, PF /œ/ is likely to be perceived as more similar to AE /u/, whereas in alveolar context, PF /œ/ is likely to be assimilated to AE /u/. The PAM, taking consonantal context into consideration, would thus predict PF /œ/ to be more difficult to differentiate from PF /u/ in alveolar context (in a single-category assimilation pattern) than in bilabial context (in a two-category assimilation pattern), a prediction supported in the present study. As listeners assimilate PF /œ/ less to AE /u/ and more to other AE vowels (e.g., /ʊ/ and /ɝ/), contrasts involving this vowel will assimilate in a two-category pattern, incurring less overlap, and discrimination accuracy is predicted to increase. However, discrimination may still be less accurate in alveo-

lar context than in bilabial, even for highly experienced L2 learners. These predictions, too, were borne out in the present study.

Consequences for speech production are addressed by Flege's (1995) SLM, which posits that when L2 segments are encountered, they are classified as "identical," "similar," or "new," relative to the listener's native phonological inventory. Viewed from this perspective, the PF vowel /u/ is classified as a similar vowel by AE speakers, resulting in inaccurate pronunciation (Flege, 1987). Flege (1987) posited that PF /y/ is categorized as a new vowel, although it might be confused with PF /u/ in the initial stages of speech learning. With L2 experience, individuals learn to distinguish PF /y/ from AE /u/, as a new category is established; thus, /y/ may be produced in a near-native manner. Results from the present study support Flege's (1987) claim in bilabial, but not in alveolar, context. The context-specific categorization patterns found in this and Levy and Strange (2008), as well as in Levy's (2009) assimilation study, suggest an allophonic level of representation in equivalence classification, wherein the consonantal context may determine whether listeners perceive a vowel as new or similar. Thus, PF /y/ may be similar to AE /u/ in alveolar context and new in bilabial context.

Perceptual training protocols that take consonantal context into consideration might better assess listeners' perceptual difficulties with vowels and gain effectiveness by targeting those contexts in which listeners have the most difficulty. These measures might help determine whether stubborn contrasts, such as PF /y-u/ for AE listeners, may ever be mastered. In training L2 learners, the cross-language assimilation overlap method may provide fruitful information for mapping the L2 sounds onto the learners' native categories in more productive ways. That is, examining the assimilation patterns that are associated with more accurate discrimination may lead to training protocols in which, for example, the similarities between PF /œ/ and AE /ɜ˞/ are emphasized to AE learners of French in the hopes that such training will help them along the steep learning curve away from /œ-u/ confusion. Furthermore, studies may ask whether perceptual training alone will result not only in improved perceptual skills, but also in more intelligible production, as has been shown in a handful of studies of Japanese listeners' perceptual training on AE consonants (Bradlow *et al.*, 1997, 1999) and AE vowels (Strange and Akahane-Yamada, 1997), as well as in perceptual training for children with phonological disorders (Rvachew, 1994). Taking into account the complex contextual variability that exists in individuals' languages is expected to result in more beneficial assimilation patterns and more accurate discrimination and comprehension as the individuals learn an L2.

ACKNOWLEDGMENTS

This research was supported by a grant to the author (NIH-NIDCD Grant No. 1F31DC006530-01). The author is indebted to Winifred Strange. Great thanks also to Allard Jongman, Geoffrey Stewart Morrison, and an anonymous reviewer for their constructive comments. Many thanks also to Loraine Obler, Martin Gitterman, Catherine Best, James Jenkins, Kanae Nishi, Valeriy Shafiro, Franzo Law II, Natalia Martínez, Gary Chant, Bruno Tagliaferri, Victoria Hatzelis, Ana de la Iglesia, and the Teachers College Speech Production and Perception Laboratory, for their helpful contributions.

APPENDIX: OVERLAP AND DISCRIMINATION SCORES

Cross-language assimilation overlap (Overlap) scores and categorical discrimination (CD) percent errors for vowel pairs in /rabVp/ and /radVt/ contexts by AE listeners with no (NoExp), moderate (ModExp), and extensive French experiences (HiExp).

	Vowel pair bp context	Overlap	CD % errors	Vowel pair dt context	Overlap	CD % errors
NoExp	/œ-i/	0	0	/œ-i/	0	0
	/u-i/	0.4	0	/œ-ɛ/	0.4	0
	/y-ɛ/	0.4	3	/y-ɛ/	0.8	0
	/œ-ɛ/	0.8	6	/u-i/	0.8	1
	/y-o/	8.5	6	/y-i/	0.9	4
	/y-i/	11.5	10	/a-ɛ/	21.3	23
	/y-u/	15.3	5	/y-o/	32.5	10
	/y-œ/	17.5	12	/y-œ/	39.3	22
	/a-ɛ/	17.5	15	/y-u/	42.3	27
	/œ-u/	36.4	31	/œ-u/	64.5	35
	/œ-o/	50.4	26	/œ-o/	65.9	39
ModExp	/u-i/	0	0	/u-i/	0	1
	/y-ɛ/	0	1	/y-ɛ/	0	1
	/œ-i/	0	3	/œ-ɛ/	0	1
	/y-i/	0.4	1	/œ-i/	0	3
	/œ-ɛ/	1.7	3	/y-i/	1.3	1
	/a-ɛ/	11.1	1	/a-ɛ/	18.4	18
	/y-o/	14.5	3	/y-o/	22.6	8
	/y-œ/	16.2	10	/y-œ/	30.4	21
	/y-u/	17.9	9	/œ-u/	33.0	26
	/œ-u/	23.1	31	/y-u/	41.8	29
	/œ-o/	24.8	15	/œ-o/	50.0	29
HiExp	/y-i/	0	0	/y-i/	0	0
	/œ-i/	0	0	/œ-i/	0	1
	/u-i/	0	0	/u-i/	0	0
	/œ-ɛ/	0	1	/y-ɛ/	0	0
	/y-ɛ/	0	1	/œ-ɛ/	0.9	1
	/a-ɛ/	0.9	0	/y-o/	1.7	4
	/y-o/	0.9	1	/a-ɛ/	2.6	4
	/œ-o/	3.0	1	/œ-o/	3.8	6
	/y-œ/	3.0	3	/y-œ/	5.6	4
	/œ-u/	13.2	1	/œ-u/	12.3	10
	/y-u/	29.1	6	/y-u/	41.9	19

[1]Front rounded PF /ø/ and /œ/ are rarely contrastive in PF. For the purposes of this paper, /œ/ represents the mid front rounded vowel.

[2]In a categorial, i.e., cross-speaker, task, the speakers differ across the three stimuli. The listeners must thus make decisions based on speaker-independent categories (Beddor and Gottfried, 1995).
[3]Except for one HiExp listener, participants in the present experiment (and in Levy, 2009) differed from those in Levy and Strange, 2008. As Levy and Strange's (2008) study had taken place 3 years prior to the present experiment, no learning effect was expected.
[4]The term "counterpart" is used loosely here to denote a speech sound for which the other language has a speech sound that is transcribed identically in broad phonetic transcription. It is acknowledged that similarly transcribed sounds may differ in their distributions of acoustic properties and that speech perception models cannot yet predict how non-native speech sounds will be mapped onto native categories (Harnsberger, 2001).
[5]A table detailing the computation of overlap is available via e-mail from the author.

Beddor, P. S., and Gottfried, T. L. (**1995**). "Methodological issues in cross-language speech perception research with adults," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 207–232.

Beddor, P. S., Harnsberger, J. D., and Lindemann, S. (**2002**). "Language-specific patters of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates," J. Phonetics **30**, 591–627.

Best, C. T. (**1995**). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 171–204.

Best, C. T., Faber, A., and Levitt, A. (**1996**). "Assimilation of non-native vowel contrasts to the American English vowel system," J. Acoust. Soc. Am. **99**, 2602.

Best, C. T., McRoberts, G. W., and Goodell, N. M. (**2001**). "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system," J. Acoust. Soc. Am. **109**, 775–794.

Best, C. T., McRoberts, G. W., and Sithole, N. M. (**1988**). "Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants," J. Exp. Psychol. **14**, 345–360.

Best, C. T., and Strange, W. (**1992**). "Effects of phonological and phonetic factors on cross-language perception of approximants," J. Phonetics **20**, 305–330.

Best, C. T., and Tyler, M. D. (**2007**). "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, edited by O.-S. Bohn and M. J. Munro (John Benjamins, Amsterdam), pp. 13–34.

Bohn, O.-S., and Steinlen, A. K. (**2003**). "Consonantal context affects cross-language perception of vowels," Proceedings of the 15th International Congress of Phonetic Sciences, pp. 2289–2292.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (**1997**). "Training Japanese listeners to identify English /r/ and /l/: Some effects of perceptual learning on speech production," J. Acoust. Soc. Am. **101**, 2299–2310.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (**1999**). "Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production," Percept. Psychophys. **61**, 977–985.

Clopper, C. G., Pisoni, D. B., and de Jong, K. (**2005**). "Acoustic characteristics of the vowel systems of six regional varieties of American English," J. Acoust. Soc. Am. **118**, 1661–1676.

Flege, J. E. (**1987**). "The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification," J. Phonetics **15**, 47–65.

Flege, J. E. (**1995**). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 233–277.

Flege, J. E., and Hillenbrand, J. (**1984**). "Limits on phonetic accuracy in foreign language speech production," J. Acoust. Soc. Am. **76**, 708–721.

Gottfried, T. L. (**1984**). "Effects of consonant context on the perception of French vowels," J. Phonetics **12**, 91–114.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (**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.

Harnsberger, J. D. (**2001**). "On the relationship between identification and discrimination of non-native nasal consonants," J. Acoust. Soc. Am. **110**, 489–503.

Hillenbrand, J. M., Clark, M. J., and Nearey, T. M. (**2001**). "Effects of consonant environment on vowel formant patterns," J. Acoust. Soc. Am. **109**, 748–763.

Iverson, P., Ekanayake, D., Hamann, S., Sennema, A., and Evans, B. G. (**2008**). "Category and perceptual interference in second-language phoneme learning: An examination of English /w/-/v/ learning by Sinhala, German, and Dutch speakers," J. Exp. Psychol. **34**, 1305–1316.

Kewley-Port, D., Burkle, T. Z., and Lee, J. H. (**2007**). "Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners," J. Acoust. Soc. Am. **122**, 2365–2375.

Kuhl, P. K., 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 (York, Timonium, MD), pp. 121–154.

Levy, E. S. (**2009**). "Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French," J. Acoust. Soc. Am. **125**, 1138–1152.

Levy, E. S., and Law, F., II (**2008**). "Production of Parisian French front rounded vowels by second-language learners," J. Acoust. Soc. Am. **123**, 3078.

Levy, E. S., and Strange, W. (**2008**). "Perception of French vowels by American English adults with and without French language experience," J. Phonetics **36**, 141–157.

Manuel, S. Y. (**1999**). "Cross-language studies: Relating language-particular patterns to other language-particular facts," in *Coarticulation: Theory, Data and Techniques*, edited by W. J. Hardcastle and N. Hewlett (Cambridge University Press, New York), pp. 179–198.

Oh, E. (**2008**). "Coarticulation in non-native speakers of English and French: An acoustic study," J. Phonetics **36**, 361–384.

Polka, L. (**1995**). "Linguistic influences in adult perception of non-native vowel contrasts," J. Acoust. Soc. Am. **97**, 1286–1296.

Polka, L., and Bohn, O.-S. (**1996**). "A cross-language comparison of vowel perception in English-learning and German-learning infants," J. Acoust. Soc. Am. **100**, 577–592.

Rochet, B. L. (**1995**). "Perception and production of second-language speech sounds by adults," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 379–410.

Rvachew, S. (**1994**). "Speech perception training can facilitate sound production learning," J. Speech Hear. Res. **37**, 347–57.

Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., and Öhman, S. (**1969**). "Cross-language study of vowel perception," Lang Speech **12**, 1–23.

Strange, W., and Akahane-Yamada, R. (**1997**). "Effects of identification training on Japanese adults' perception of American English vowels," J. Acoust. Soc. Am. **102**, 3137.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., and Nishi, K. (**2001**). "Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners," J. Acoust. Soc. Am. **109**, 1691–1704.

Strange, W., Bohn, O.-S., Nishi, K., and Trent, S. A. (**2005**). "Contextual variation in the acoustic and perceptual similarity of North German and American English vowels," J. Acoust. Soc. Am. **118**, 1751–1762.

Strange, W., Levy, E. S., and Law, F. F., II (**2009**)."Cross-language categorization of French and German vowels by naïve American listeners," J. Acoust. Soc. Am. **126**(3), 1461–1476.

Strange, W., and Shafer, V. L. (**2008**). "Speech perception in second language learners: The reeducation of selective perception," in *Phonology and Second Language Acquisition*, edited by J. G. Hansen Edwards and M. L. Zampini (John Benjamins, Amsterdam), pp. 153–191.

Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., and Nishi, K. (**2007**). "Acoustic variability within and across German, French and American English vowels: Phonetic context effects," J. Acoust. Soc. Am. **122**, 1111–1129.

Tranel, B. (**1987**). *The Sounds of French: An Introduction* (Cambridge University Press, New York).

Consonant recognition loss in hearing impaired listeners

Sandeep A. Phatak
Army Audiology and Speech Center, Walter Reed Army Medical Center, Washington, DC 20307

Yang-soo Yoon
House Ear Institute, 2100 West 3rd Street, Los Angeles, California 90057

David M. Gooler
Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign, Urbana, Illinois 61820

Jont B. Allen
ECE Department and the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

(Received 12 September 2008; revised 9 July 2009; accepted 1 September 2009)

This paper presents a compact graphical method for comparing the performance of individual hearing impaired (HI) listeners with that of an average normal hearing (NH) listener on a consonant-by-consonant basis. This representation, named the *consonant loss profile* (CLP), characterizes the effect of a listener's hearing loss on each consonant over a range of performance. The CLP shows that the consonant loss, which is the signal-to-noise ratio (SNR) difference at equal NH and HI scores, is consonant-dependent and varies with the score. This variation in the consonant loss reveals that hearing loss renders some consonants unintelligible, while it reduces noise-robustness of some other consonants. The conventional SNR-loss metric ΔSNR_{50}, defined as the SNR difference at 50% recognition score, is insufficient to capture this variation. The ΔSNR_{50} value is on average 12 dB lower when measured with sentences using standard clinical procedures than when measured with nonsense syllables. A listener with symmetric hearing loss may not have identical CLPs for both ears. Some consonant confusions by HI listeners are influenced by the high-frequency hearing loss even at a presentation level as high as 85 dB sound pressure level.
 [DOI: 10.1121/1.3238257]

PACS number(s): 43.71.Ky, 43.71.Es [MSS] Pages: 2683–2694

I. INTRODUCTION

Consonant recognition studies have shown that hearing impaired (HI) listeners make significantly more consonant errors than normal hearing (NH) listeners both in quiet (Walden and Montgomery, 1975; Bilger and Wang, 1976; Doyle *et al.*, 1981) and in presence of a noise masker (Dubno *et al.*, 1982; Gordon-Salant, 1985). It has been shown that NH listeners demonstrate a wide range of performance across different consonants presented in noise (Phatak and Allen, 2007; Phatak *et al.*, 2008), and it is likely that HI listeners would also exhibit a variance in performance across the same consonants. A quantitative comparison of recognition performances for individual consonants is necessary to determine whether the loss of performance for HI listeners is consonant-dependent. In this study, we present a graphical method to quantitatively compare the individual consonant recognition performance of HI and NH listeners over a range of signal-to-noise ratios (SNRs). Such comparison characterizes the impact of hearing loss on the perception of each consonant.

The effect of hearing loss on speech perception in noise has two components. The first component is the loss of audibility, which causes elevated thresholds for any kind of external sound. Modern hearing aids compensate this loss with a multichannel non-linear amplification. In spite of the amplification, HI listeners still have difficulty perceiving speech in noise compared to NH listeners (Humes, 2007). This difficulty in understanding speech in noise at supra-threshold sound levels is the second component and is known by different names such as "distortion" (Plomp, 1978), "clarity loss," and "SNR-loss" (Killion, 1997). Audibility depends on the presentation level of the sound whereas the supra-threshold performance depends on the SNR. It is difficult to separate the two components, and this has led to controversies about characterizing of the SNR-loss. Technically, SNR-loss is defined as the additional SNR required by a HI listener to achieve the same performance as a NH listener. The difference between *speech reception thresholds* (SRTs) of HI and NH listeners, at loud enough presentation levels, is the commonly used quantitative measure of SNR-loss (Plomp, 1978; Lee and Humes, 1993; Killion, 1997). The SRT is the SNR required for achieving 50% recognition performance. We will use ΔSNR_{50} to refer to this metric, and the phrase SNR-loss to refer to the phenomenon of SNR deficit exhibited by HI listeners at supra-threshold levels.

Currently, the most common clinical method for measuring SNR-loss is the QuickSIN test developed by Etymotic Research (Killion *et al.*, 2004). In this procedure, two lists of six IEEE sentences are presented at about 80–85 dB sound pressure level (SPL)[1] to listeners with pure-tone average (PTA) (i.e., the average of audiometric thresholds at 500 Hz,

TABLE I. Listener information: subject ID, ear tested (L: left, R: right), age (in years), gender (M: male, F: female), average pure-tone thresholds at 0.25, 0.5, 1, and 2 kHz (PTA_{LF}) and at 4 and 8 kHz (PTA_{HF}), and the listener set (LS). Six listeners were common to both sets LS1 and LS2.

Sub. ID	1	2	3	4	12	39	48	58	71	76	112	113	134	148	170	177	188	195	200	208	216	300	301
Ear	L	L,R	R	L,R	L	L	R	R	L	L	R	R	L	L	R	R	R	L	L,R	L	L	L	R
Age	21	63	21	59	39	63	62	55	60	62	54	48	52	60	53	39	64	60	52	54	58	54	58
Gender	F	F	M	F	F	M	M	F	M	F	F	M	F	M	M	F	M	F	M	F	F	M	F
PTA_{LF}	27.5	L:6.25 R:10	26.25	L:40 R:38.75	31.25	26.25	26.25	43.75	25	42.5	11.25	42.5	18.75	15	3.75	31.25	16.25	16.25	L:20 R:25	21.25	50	8.75	17.5
PTA_{HF}	85	L:60.0 R:62.5	40	L:52.5 R:47.5	82.5	60	70	22.5	60	55	67.5	60	52.5	67.5	35	52.5	45	52.5	L:55.0 R:57.5	47.5	65	50	90
LS	1	1	1	1	1,2	1	1,2	1,2	1	1	1,2	1,2	1,2	1	2	2	2	2	2	2	2	2	2

1 kHz, and 2 kHz) up to 45 dB hearing level (HL) and at a "Loud, but OK" level to those with PTA $\geq$ 50 dB HL (QuickSIN manual, 2006) without pre-emphasis or high-frequency amplification. The first sentence of each list is presented at 25 dB SNR with a four-talker babble masker, and the SNR of subsequent sentences is decreased in 5 dB steps. The SRT is then estimated using method of Hudgins *et al.* (1947) described in Tillman and Olsen (1973), and ΔSNR_{50} is calculated by assuming that the average SRT for NH listeners on the same task is 2 dB SNR. The ΔSNR_{50} is presumed to be a measure of supra-threshold SNR-loss. However, the QuickSIN procedure does not guarantee that the HI listener listens at supra-threshold levels at all frequencies. Therefore, the ΔSNR_{50} thus obtained may not necessarily be an audibility-independent measure of the SNR-loss. Also, ΔSNR_{50} does not provide any information about the SNR differences at other performance points, e.g., 20% or 80% recognition, and therefore it is insufficient to characterize the SNR-loss.

Another difficulty in interpreting the results of most SNR-loss studies is that the speech stimuli used are either sentences or meaningful words (spondee). A context effect due to meaning, grammar, prosody, etc., increases the recognition scores in noisy conditions without actually improving the perception of speech sounds (French and Steinberg, 1947; Boothroyd and Nittrouer, 1988). The effect of context information on the ΔSNR_{50} metric has not been measured.

To address such issues regarding SNR-loss and consonant identification, a Miller and Nicely (1955) type confusion matrix (CM) experiment was conducted on HI listeners. The results of this experiment are compared with NH data from Phatak and Allen (2007) collected on the same set of consonant-vowel (CV) stimuli. In this paper, we develop a compact graphical representation of such comparison for each HI listener on a consonant-by-consonant basis. This representation, denoted the *consonant loss profile* (CLP), shows variations in the individual consonant loss over a range of performance. We also compare consonant confusions of HI and NH listeners. To estimate the effect of context, ΔSNR_{50} values obtained from our consonant recognition experiment will be compared with those obtained using the QuickSIN.

II. METHODS

The listeners in this study were screened based on their pre-existing audiograms and other medical history. Anyone with a conductive hearing loss or a history of ear surgery was excluded. All those listeners having sensorineural hearing loss with PTA between 30 and 70 dB HL, in at least one ear, were recruited for this study and their audiograms were re-measured. Table I shows details of the participants. Initially, 12 listeners were tested with their best ears, while both ears of 2 listeners were separately tested, resulting in 16 HI test ears. We call these as listener set 1 (LS1). 1 year later, six of the LS1 HI ears along with nine new HI ears (eight new HI listeners) were tested (LS2). In all, 26 HI ears were tested in this study.

The degree of audiometric hearing loss of listeners varied from normal to moderate at frequencies up to 2 kHz ($PTA_{LF} \leq 55$ dB HL) and between moderate to profound at 4 kHz and above (PTA_{HF}). The QuickSIN thresholds (standard, no filtering) were also measured for LS1 ears (Killion *et al.*, 2004). Listeners were tested unaided in a sound-treated room at the Speech and Hearing Science Department of the University of Illinois at Urbana-Champaign. The testing procedures for the CM experiment were similar to those used in Phatak and Allen's (2007) study. Isolated CV syllables with 16 consonants (/p/, /t/, /k/, /f/, /θ/, /s/, /ʃ/, /b/, /d/, /g/, /v/, /ð/, /z/, /ʒ/, /m/, /n/) and vowel /ɑ/, each spoken by ten different talkers, were selected from LDC2205S222 [2] database (Fousek *et al.*, 2004). The stimuli were digitally recorded at a sampling rate of 16 kHz. The syllables were presented in noise masker at five SNRs (−12, −6, 0, 6, 12 dB) and in the quiet condition (i.e., no masker). The masker was a steady-state noise with an average speech-like spectrum identical to that used by Phatak and Allen (2007). The masker spectrum was steady between 100 Hz and 1 kHz and had low-frequency and high-frequency roll-offs of 12 and −30 dB/dec, respectively. Listeners were tested monaurally using Etymotic ER-2 insert earphones to avoid ear canal collapse and to minimize cross-ear listening. The presentation level of the clean speech (i.e., the quiet condition) was adjusted to the most comfortable level (MCL) for each listener using an external TDT PA5 attenuator. The attenuator setting was maintained for that listener throughout the experiment. Though the actual speech level for each listener was not measured, calibration estimates the presentation levels in the ear canal to be either 75 or 85 dB SPL depending on the attenuator setting. Each token was level-normalized before presentation using VU-METER software (Lobdell and Allen, 2007). No filtering was applied to the stimuli.

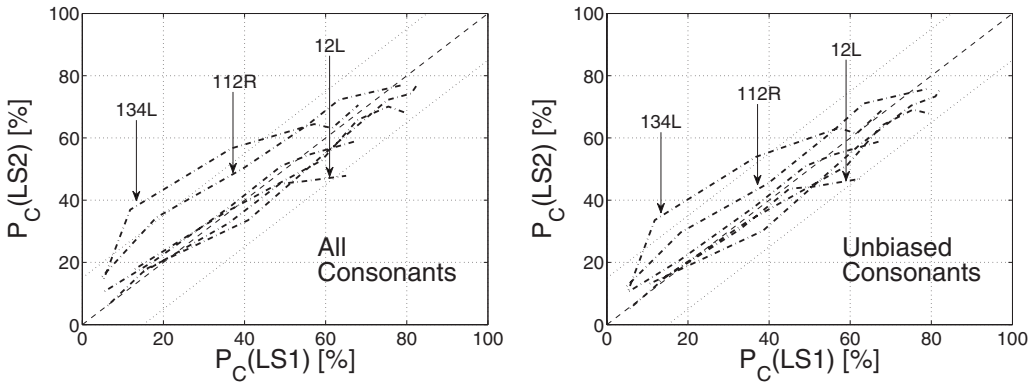


FIG. 1. (a) The average consonant scores in listener sets 1 [P_c(LS1)] and 2 [P_c(LS2)] for the six common listeners 12L, 48R, 76L, 112R, 113R, and 134L. (b) Same as (a), but with average consonant scores estimated using only unbiased consonants, as discussed in the text. The dotted lines parallel to diagonal represent ±15% difference ($|\Delta P_c| = |P_c(\mathrm{LS2}) - P_c(\mathrm{LS1})| = 15\%$).

Listeners were asked to identify the consonant in the presented CV syllable by selecting 1 of 16 software buttons on a computer screen, each labeled with one consonant sound. A pronunciation key for each consonant was provided next to its button. This was necessary to avoid confusions due to orthographic similarity between consonants such as /θ/-/ð/ and /z/-/ʒ/. The listeners were allowed to hear the syllable as many times as they desired before making a decision. After they clicked their response, the next syllable was presented after a short pause. The syllable presentation was randomized over consonants, talkers, and SNRs.

The performance of each HI listener is then compared with the average consonant recognition performance of ten NH listeners. These NH data are a subset of the Phatak and Allen (2007) data (16 consonants, 4 vowels) that includes responses to the CV syllables with only vowel /ɑ/, presented in speech-weighted noise (SNR: [−22, −20, −16, −10, −2] dB and quiet). In Phatak and Allen's (2007) study, the stimuli were presented diotically via circumaural headphones. Though diotic presentation increases the percept of loudness, it does not provide any significant SNR-advantage over monaural presentation for recognizing speech in noise (Licklider, 1948; Helfer, 1994). Therefore, the results of this study do not require any SNR-correction for comparing with the results of Phatak and Allen (2007).

III. RESULTS

A. Test-retest measure

There was a gap of about 1 year between data collection from the two listener sets LS1 and LS2. The scores of six HI listeners, which were common to both sets, are compared for consistency in Fig. 1(a). The average consonant recognition score P_c(SNR) is the ratio of the number of consonants recognized correctly to the total number consonants presented at a given SNR. Three of the six listeners, viz., 48R, 76L, and 113R, have very similar scores in both sets with correlation coefficients (r) greater than 0.99 (Table II). Listeners 12L, 112R, and 134L have score differences of more than 15% across the two sets (i.e., $|\Delta P_c| = |P_c(\mathrm{LS2}) - P_c(\mathrm{LS1})| > 15\%$). However, these differences are primarily due to two to four consonants. Confusion matrices (not shown) reveal that these three listeners have significant biases toward specific consonants in the LS2 data. We define a listener to have a bias in favor of a consonant if at the lowest SNR (i.e., −12 dB), the total responses for that consonant are at least three times the number of presentations of that consonant. Such biases result in high scores. For example, at −12 dB SNR, listener 134L has a /s/ recognition score of 70% due to a response bias indicated by a high false alarm rate for /s/ (i.e., 27 presentations of consonant /s/ and 87 /s/ responses, out of which 19 were correct). Listener 76L also showed bias for consonant /s/, whereas listeners 112R showed bias for /t/ and /z/ in LS2 data. Listeners 12L and 113R showed biases for /n/ and /s/, respectively, but in both sets. Such biases, though present, are relatively weaker in set LS1 for these three listeners. If average scores in both sets are estimated using only unbiased consonants for each listener, then five out of six listeners have score differences less than 15% [i.e., $|\Delta P_c| \leq 0.15$, Fig. 1(b)]. The higher LS2 scores for listener 134L could be a learning effect. Therefore, only LS1 data were used for 134L. For the other five listeners, both LS1 and LS2 data were pooled together before estimating CMs.

B. Context effect on ΔSNR_{50}

The consonant recognition SRT for each HI listener was obtained by interpolating the P_c(SNR) values for that listener. The ΔSNR_{50} was then calculated by comparing this SRT value to the corresponding SRT for average NH (ANH) performance from Phatak and Allen (2007). The ΔSNR_{50} values thus obtained for consonant recognition are compared with the measured QuickSIN ΔSNR_{50} values for LS1 listeners in Fig. 2. The two ΔSNR_{50} values are not statistically correlated ($p > 0.05$) with each other. The variability across

TABLE II. The correlation coefficients (r) between P_c(LS1) and P_c(LS2) scores for the six listeners from Fig. 1.

Listener	12L	48R	76L	112R	113R	134L
All consonants	0.980	0.992	0.992	0.987	0.994	0.946
Unbiased consonants	0.978	0.992	0.994	0.990	0.994	0.951

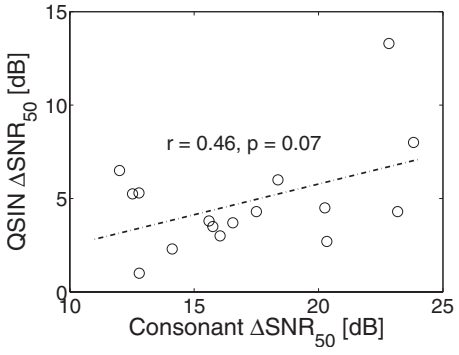


FIG. 2. A comparison of ΔSNR_{50} values estimated from our consonant recognition data (abscissa) and those measured using QuickSIN (ordinate). A linear regression line (dashed-dotted) is shown along with the correlation coefficient (r) and null-hypothesis probability (p) for the two sets of ΔSNR_{50} values.

listeners is greater for consonant ΔSNR_{50} than for the QuickSIN ΔSNR_{50}. Consonant ΔSNR_{50} values are moderately correlated ($p<0.03$) with the average of pure-tone thresholds at 4 and 8 kHz (PTA_{HF}), accounting for 30% of the variance ($r=0.55$) in consonant ΔSNR_{50}. The QuickSIN ΔSNR_{50} values are neither correlated with PTA_{HF} nor with the traditional PTA (0.5, 1, and 2 kHz).

The sentence ΔSNR_{50} obtained using QuickSIN is, on average, 12.31 dB lower than the consonant ΔSNR_{50}. The difference in SRT estimation procedures [i.e., Tillman and Olsen (1973) vs linear interpolation of P_c(SNR)] cannot account for such a large difference. Therefore, the most likely reason for this difference could be the speech stimuli. Sentences have a variety of linguistic context cues that are not present in isolated syllables. HI listeners rely more than NH listeners on such context information in order to compensate for their hearing loss (Pichora-Fuller *et al.*, 1995). This results in lower ΔSNR_{50} values for sentences than for consonants. The context effect may also be responsible for the absence of correlation between QuickSIN ΔSNR_{50} and the PTA values.

C. Average consonant loss

The average consonant recognition errors $P_e(\text{SNR})=1-P_c(\text{SNR})$ for HI listeners are significantly higher than those for NH listeners. Figure 3(a) shows P_e(SNR) for three different HI listeners (3R, 58R, and 113R) and for the ANH data (solid line) on a logarithmic scale. The conventional ΔSNR_{50} measure would report the SNR difference between the 50% points (i.e., the SRT difference) of ANH and HI listeners denoted by the circles. The ANH and HI P_e(SNR) curves are not parallel, and therefore the SNR difference ΔSNR is a function of the performance level P_e. For example, listeners 3R and 113R have an identical SNR difference at $P_e=50\%$, but listener 113R has a greater SNR difference than 3R at $P_e=20\%$. Listener 58R cannot even achieve an error of 30%. Since the performance-SNR functions of HI and NH listeners are not parallel, the SNR difference at a single performance point, such as the SRT, is not sufficient to characterize the SNR deficit.

To characterize the SNR difference over a range of performance, we plot the SNR required by a HI listener against that required by an average NH listener to achieve the same average consonant score. We call this matched-performance SNR contour as the *average consonant loss* curve. Figure 3(b) shows the average consonant loss curves for 3R, 58R, and 113R obtained by comparing their scores with the ANH scores from Fig. 3(a). The dash-dotted straight lines show contours for SNR differences ($\Delta\text{SNR}=\text{SNR}_{HI}-\text{SNR}_{ANH}$) of 0, 10, and 30 dB. In this example, 58R achieves 50% performance (circle) at 2 dB SNR, while the 50% ANH score is achieved at −16 dB SNR. Thus, the 50% point on the average consonant loss curve [circle in Fig. 3(b)] for 58R has an abscissa of $\text{SNR}_{HI}=2$ dB and an ordinate of $\text{SNR}_{ANH}=-16$ dB. The conventional $\Delta SNR_{50}=(\text{SNR}_{HI}-\text{SNR}_{ANH})|_{P_c=50\%}$ value can be obtained for consonants by measuring the distance of this point from the $\Delta\text{SNR}=0$ line, which depicts zero consonant loss (i.e., identical ANH and HI SNRs for a given performance). A curve below this line indicates that the HI listener performance is worse than the ANH performance (i.e., $\Delta\text{SNR}>0$). These average conso-

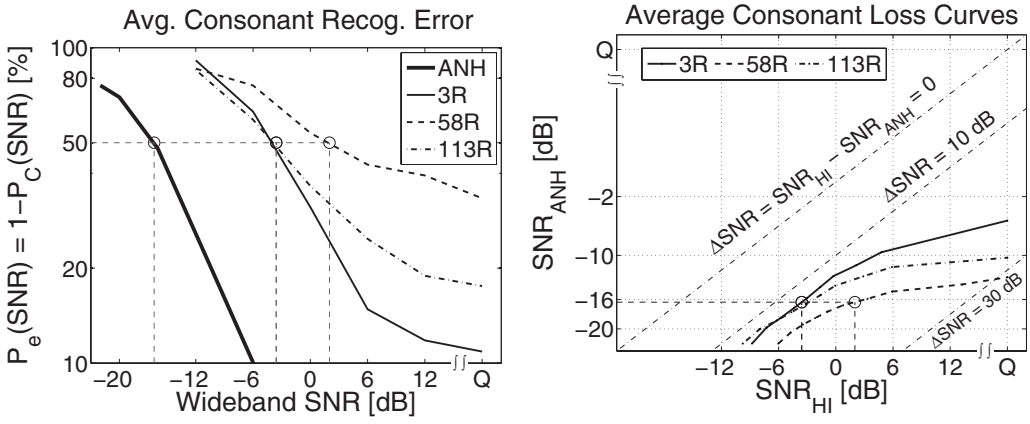


FIG. 3. (a) Average consonant recognition error P_e(SNR) for three HI listeners 3R, 58R, and 113R. The thick solid line denotes the corresponding P_e(SNR) for the average of ten NH listener data (i.e., ANH) from Phatak and Allen (2007). The circles represent the 50% performance points. (b) Average consonant loss curves for the same three listeners.

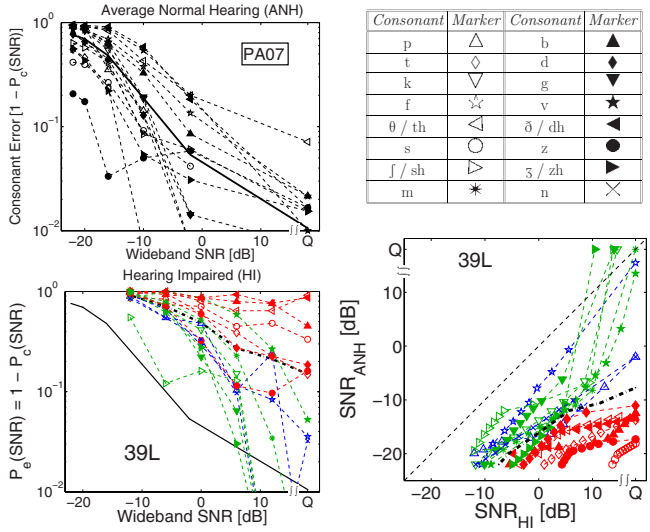


FIG. 4. (Color online) Top left: Consonant recognition errors P_e(SNR) on a log scale for average normal hearing data from Phatak and Allen (2007). The solid line shows the average consonant error. Top right: Marker symbols for the consonants. Bottom left: P_e(SNR) for HI listener 39, left ear. Thick, dash-dotted line shows the average consonant error for 39L, while solid line shows average consonant error for the ANH data (from top left panel). Bottom right: The consonant loss profile for HI ear 39L. The dashed reference line with no symbols represents $\Delta\text{SNR}=\text{SNR}_{\text{HI}}-\text{SNR}_{\text{ANH}}=0$ and the thick dash-dotted line is the average consonant loss curve.

nant loss curves diverge from the $\Delta\text{SNR}=0$ reference line in the $\text{SNR}_{\text{HI}}>0$ dB region, indicating a greater consonant loss. Such variation in consonant loss cannot be characterized by the ΔSNR_{50} metric.

D. Consonant loss profile

There is a significant variation in the performance of individual HI listeners across different consonants. The bottom left panel of Fig. 4 shows consonant error P_e(SNR) for HI ear 39L (i.e., listener 39, left ear) on a log scale. At 0 dB SNR, the average consonant error is 51%, but the error varies across consonants from 16% for /ʃ/ to 87% for /v/, resulting in a standard deviation of $\sigma_{P_e(0\text{ dB})}=24\%$. Similar large variations in consonant errors are observed for all HI ears with $\sigma_{P_e(0\text{ dB})}$ ranging from 19% to 34%. For the ANH consonant errors (Fig. 4, top left panel) the average error at −16 dB is 52%, but individual consonant error varies from 3% to 90% with $\sigma_{P_e(-16\text{ dB})}=25\%$. Thus, a comparison of the average curves for ANH and HI listeners does not provide useful information about the SNR deficits for individual consonants.

To analyze the loss for individual consonants, the matched-performance SNR contours are plotted for each consonant for a given HI ear. We call this plot the *consonant loss profile* (CLP) of that ear. The bottom right panel in Fig. 4 shows the CLP for 39L. Each of the 16 CLP curves has ten data points, indicated by the marker positions, but with different spacing. For estimating the consonant loss curve for a consonant, first a range of performance P_e that was common to both HI and ANH performances for that consonant was determined and then SNR_{HI} and SNR_{ANH} values for ten equidistant P_e points in this range were estimated by interpolating P_e(SNR) curves, resulting in different marker spacings on SNR scale. Thus, some consonants that have a small common P_e range across HI and ANH performance, such as /s/ (○), have closely spaced markers, while others like /f/ (☆) with a larger common P_e range have widely spaced markers.

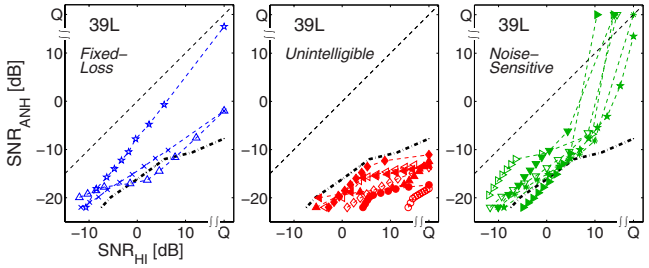


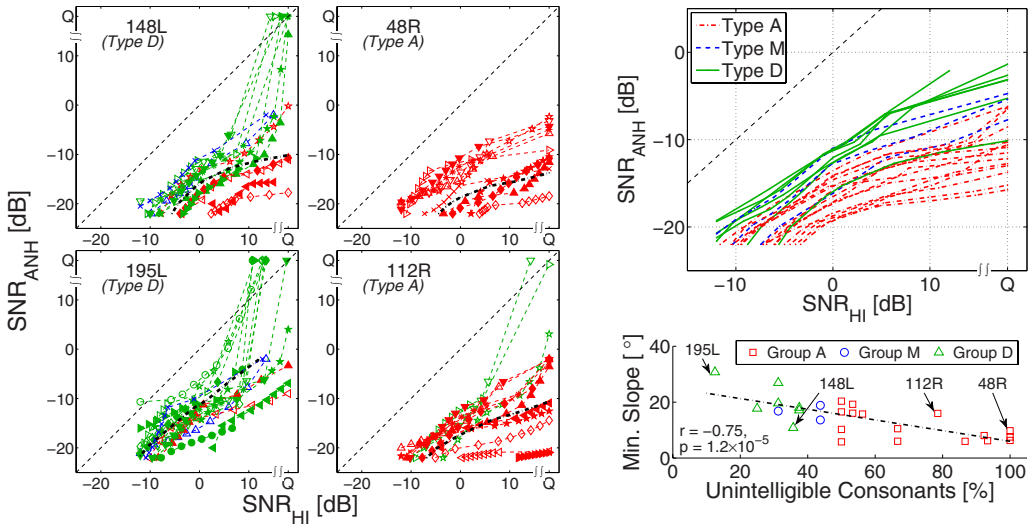
FIG. 5. (Color online) The consonant-loss profile curves of 39L separated into three groups: *fixed-loss* (/p/,/f/,/n/) (left), *unintelligible* (/t/,/θ/,/s/,/b/,/d/,/ð/,/z/) (center), and *noise-sensitive* (/k/,/ʃ/,/g/,/ʒ/,/v/,/m/) (right).

These individual consonant loss curves have patterns that are significantly different than the average curve (thick dash-dotted curve). The curves can be categorized into three sets, as shown in Fig. 5.

(1) *Fixed-loss* consonants. These curves run almost parallel to the reference line with no significant slope change. The SNR difference for these consonants is nearly constant.
(2) *Unintelligible* consonants. These curves, with very shallow slopes, indicate poor intelligibility even at higher SNR for HI listeners. The quiet condition (Q) performance of the HI ear for these consonants $[P_c(Q)\approx 50\%]$ is equivalent to the ANH performance at about −10 to −12 dB SNR. These consonants clearly indicate an audibility loss even at the MCL.
(3) *Noise-sensitive* consonants. The HI ear performance for these consonants is close to, and sometimes even better than, the ANH performance in the quiet condition. In quiet, these consonants are not affected by the elevated thresholds of the HI ear. However, even a small amount of noise (i.e., $\text{SNR}_{\text{HI}}=12$ dB) reduces the HI ear's performance to an ANH equivalent of −8 to −12 dB SNR, resulting in significant SNR deficits for these noise-sensitive consonants.

The CLP characterizes the type and magnitude of the effect of hearing loss on each consonant. The unintelligible and the noise-sensitive consonant groups relate to Plomp's *A*-factor (audibility) and *D*-factor (distortion) losses, respectively. For the purpose of this study, consonant loss curves that do not exceed a slope (i.e., $\Delta\text{SNR}_{\text{ANH}}/\Delta\text{SNR}_{\text{HI}}$) of $\tan(30°)=0.577$ dB/dB are categorized as unintelligible, while those which achieve a slope greater than $\tan(60°)=1.732$ dB/dB are categorized as noise-sensitive.

Note that the CLP is a relative metric because it is referenced to the ANH consonant scores. A consonant with the highest error, such as /θ/ (◁) or /ð/ (◀) for 39L (Fig. 4, bottom left panel), may not have the highest SNR difference (Fig. 4, bottom right panel) if the ANH performance for that consonant is also poor.



(a) (b)

FIG. 6. (Color online) (a) Examples of type A (148L, 195L) and type D (48R, 112R) consonant loss profiles. The average consonant loss curves (thick, dash-dotted) are shown in black. (b) Top: The average consonant loss curves for all HI listeners. The CLP types associated with the curves are represented by line styles. Bottom: Comparison of percentage of unintelligible sounds in each CLP vs the minimum slope (in degrees) of the corresponding average consonant loss curve. Regression line and correlation coefficient are shown.

1. CLP types

The degree and type of consonant loss for a given consonant vary significantly across listeners. Every HI listener shows a different distribution of consonants in each of the three CLP sets (i.e., fixed-loss, noise-sensitive, and unintelligible). Depending on the number of consonants in each set, we divide the HI CLPs into three types. If more than half the consonants in a CLP are unintelligible (i.e., *A*-factor loss), then it is categorized as type A CLP. If more than half the consonants are noise-sensitive (i.e., *D*-factor loss), then the CLP is type D. All other CLPs, with intermediate distributions, are labeled as type M.

Figure 6(a) shows two type A and two type D CLPs. The average curves (dash-dotted) for type A CLP have shallower slopes than for type D due to higher percentage of unintelligible consonants. Also, within each type, more unintelligible consonants result in a flatter average curve. In general, the number of unintelligible consonants determines the shape of average consonant loss curves, i.e., a CLP with more unintelligible consonant curves has a flatter average consonant loss curve. The bottom panel in Fig. 6(b) shows that the minimum slope of average consonant loss curve is inversely proportional ($r=-0.75$, $p<0.0001$) to the percentage of unintelligible consonants. The top panel in Fig. 6(b) shows the average consonant loss curves for the 26 HI ears tested. The minimum slopes are always in the $SNR_{HI}>0$ region. Though the curves associated with type A CLPs are relatively flatter than those associated with the other two CLP types, there is no well-defined boundary. These curves rather show a near continuum of average consonant loss.

Figure 7 shows audiograms for all HI ears tested. The distribution of consonants in the three CLP sets and the listener group based on that distribution do not correlate well with the audiograms. Listeners with type A CLP (dash-dotted lines) have relatively higher hearing loss than the rest of listeners, but this difference is not statistically significant. Also, one type A listener (58R) has very good high-frequency hearing (15 dB HL at 8 kHz). This outlier listener has moderate hearing loss only in the mid-frequency range (1–2 kHz), suggesting the possibility of a cochlear dead region.

2. Ear differences

Both ears were tested for three HI listeners (2, 4, and 200). The left and right ear CLPs for two of these listeners (2 and 4) are compared in Fig. 8. Both listeners have symmetric hearing losses (right panel), and there are many similarities between the two CLPs of each listener. There are also some significant differences. For example, consonant /ʃ/ (▷) is noise-sensitive for the right ear of listener 2, but not for the left ear. Similarly consonant /ð/ (◀) is unintelligible to listener 4's left ear, but not to the right ear. These differences between ears imply differences in the peripheral auditory system that are not accounted for by only the audiograms.

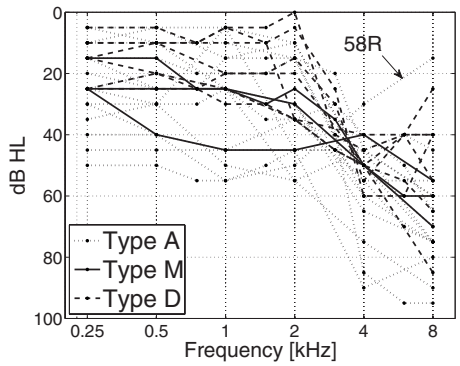


FIG. 7. Audiograms of the HI ears. The CLP type associated with each ear is represented by the line style.

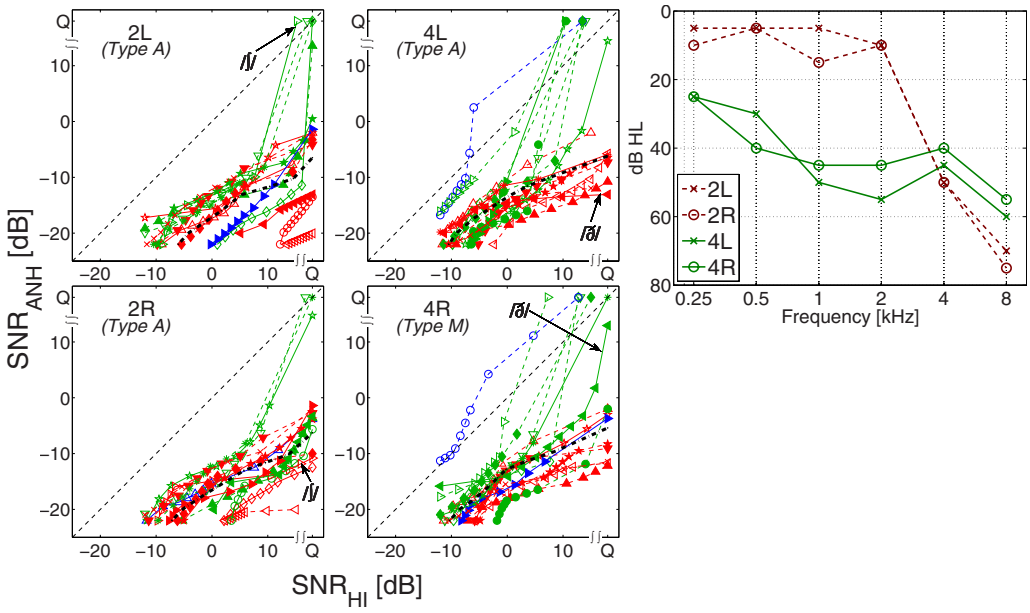


FIG. 8. (Color online) Consonant loss profiles for both ears of two listeners: 2L, 2R (left) and 4L, 4R (center). For each listener, the consonants that are categorized differently in two ears are shown with solid curves. Audiograms for the four ears are shown in the right panel.

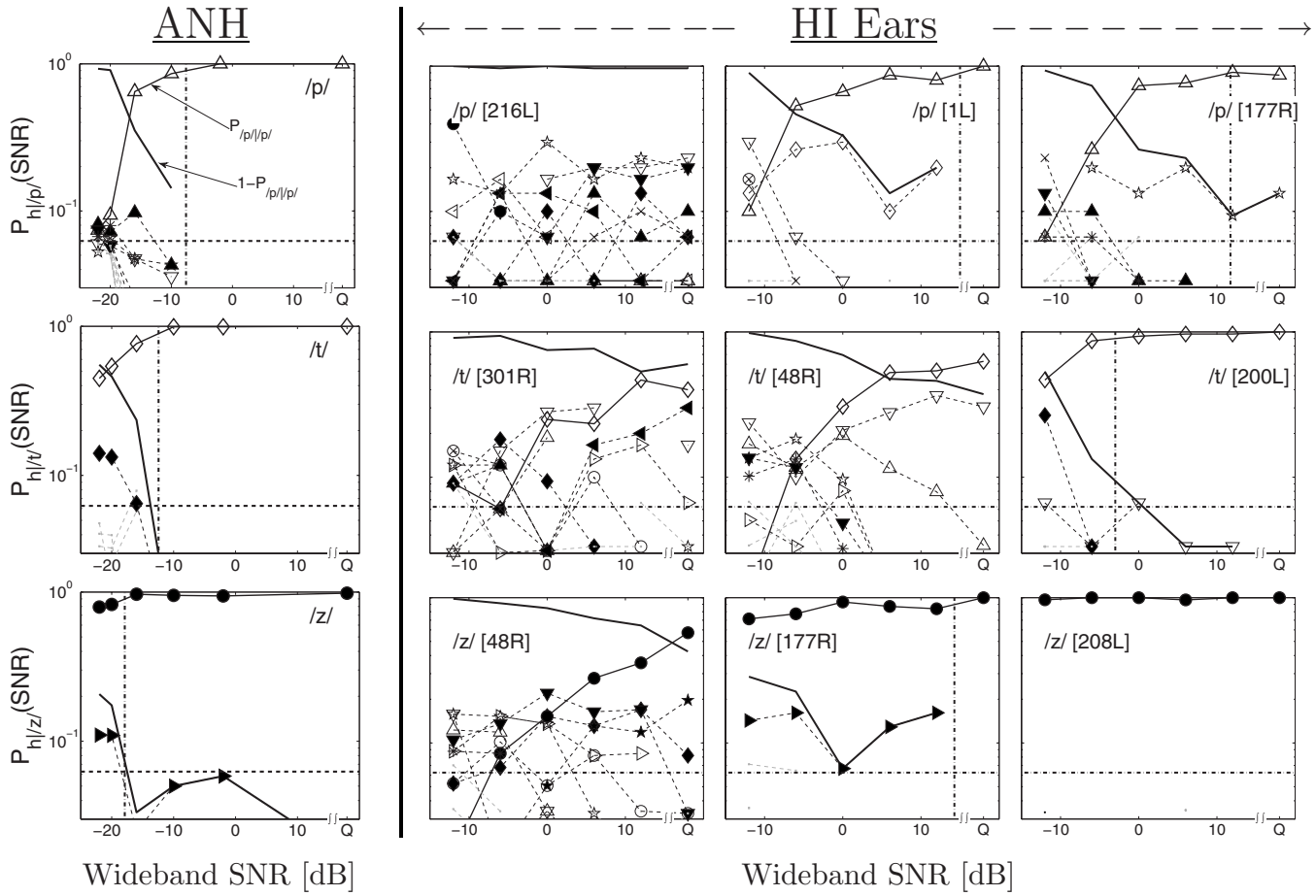


FIG. 9. Three examples of consonant CPs for the ANH data are shown in the left column. The three examples are for consonants /p/ (Δ, top row), /t/ (◇, center row), and /z/ (●, bottom row). The consonant CPs from three different HI ears (right three columns) are shown for each of these three consonants. Each dashed curve represents confusion with a consonant, denoted by the symbol. The legend of consonant symbols is provided in Fig. 4. The vertical dashed line in each panel indicates SNR_{90}, i.e., the SNR at P_c=90%.

E. Consonant confusions

We use *confusion patterns* (CPs) to analyze consonant confusions. A confusion pattern [$P_{h|s}$(SNR); *s*: spoken, *h*: heard] is a plot of all the elements in a row of a row-normalized CM (i.e., row sum=1) as a function of SNR. For example, the top left panel in Fig. 9 shows the row corresponding to the presentation of consonant /p/ in the ANH CM [$P_{h|s}$(SNR) for $s=$/p/]. Each curve with symbols shows values of one cell in the row as a function of SNR. For example, the curve $P_{/p/|/p/}$(SNR) represents the diagonal cell, i.e., the recognition score for /p/. The solid line without any symbols represents the total error, i.e., $1-P_{/p/|/p/}$(SNR). All other dashed curves are off-diagonal cells, which represent confusions of /p/ with other consonants, denoted by the symbols. The legend of consonant symbols is shown in Fig. 4. The confusion of /p/ with /b/ (▲ in top left panel), a voicing confusion, has a maximum at −16 dB SNR. Thus, consonant /p/ forms a weak confusion group with /b/, below the *confusion threshold* of −16 dB SNR. At very low SNRs, all curves asymptotically converge to the chance performance of 1/16 (horizontal dashed line).

The ANH CPs have a small number of confusions, but with clearly formed maxima, for all consonants. For example, /t/ (◇, center left panel) forms a confusion group with /d/ (◆), and /z/ (●, bottom left panel) with /ʒ/ (▶). While this is true for the majority of HI CPs, a few HI CPs show several simultaneous confusions with poorly defined confusion groups. One such HI CP is shown in Fig. 9 for each of the three exemplary consonants (216L for /p/, 301R for /t/, and 48R for /z/). The HI ears that exhibit CPs with a large number of competitors are different across consonants.

Figure 9 also shows two HI CPs with a small number of competitors (panels in columns 3 and 4) for each of the three consonants. Clear consonant confusion groups are observed in these CPs due to a small number of competitors. For some of these CPs, the confusion groups are similar to those observed in the ANH data. For example, /t/-/d/ confusion for 200L and /z/-/ʒ/ confusion for 177R are also observed in the corresponding ANH CPs (left column). However, several HI CPs show confusion groups that are different from those in the corresponding ANH CPs. For example, the ANH CP for /p/ [i.e., $p_{h|/p/}$(SNR)] shows a weak /p/-/b/ group, while HI ears show significant confusions of /p/ with /t/ and /k/ (1L) and with /f/ (177R). Similarly, HI listeners often confuse /t/ with /p/ and /k/ (48R), unlike the /t/-/d/ confusion in the ANH CP (left column, center panel). Occasionally, a HI listener may show better performance than ANH. For example, 208L (bottom right panel) did not confuse /z/ with any other consonant, resulting in a score >90% even at a SNR of −12 dB.

The /t/-/p/-/k/ confusion group from the HI CPs (/p/ for 1L and /t/ for 48R) is not observed in the corresponding ANH data (left column, top 2 panels). However, NH listeners show the same confusion group in the presence of a white noise masker (Miller and Nicely, 1955; Phatak *et al.*, 2008). This is because white noise masks higher frequencies more than speech-weighted noise at a given SNR. Several studies have shown that the bandwidth and intensity of the release burst at these high frequencies are crucial when distinguishing stop plosive consonants (Cooper *et al.*, 1952; Régnier and Allen, 2008). In HI ears, audiometric loss masks these high frequencies, resulting in confusions similar to those for NH listeners in white noise. Note that 200L (PTA_{HF} =55 dB HL), with relatively better high-frequency hearing than 48R (PTA_{HF}=70 dB HL), did not show /t/-/p/-/k/ confusion.

HI CPs are not as smooth as the ANH CPs. This is because the ANH data are pooled over many listeners, thereby increasing the row sums of CMs and decreasing the variance. The HI data are difficult to pool over listeners because of the large confusion heterogeneity across HI ears compared to NH data. The low row sums for listener-specific HI CMs increase the statistical variance of the confusion analysis, especially when there are multiple competitors. In such cases, the probability distribution of a row becomes multimodal, which is difficult to estimate reliably with low sample size (i.e., row sum). When the distribution is unimodal (no confusions) or bi-modal (one strong competitor), low row sums are adequate for obtaining reliable confusion patterns.

The HI listeners not only make significantly more errors than the ANH errors but also demonstrate *different* errors. In terms of the CPs, it means that the HI CPs not only have higher confusion thresholds but also have different competitors than those in the corresponding ANH CPs. Phatak *et al.* (2008) described two characteristics of the variability in confusion patterns. First is *threshold variability*, where the SNR below which the recognition score of a consonant drops sharply, varies across CPs. The second type of variability is *confusion heterogeneity*, where competitors of the same sound are different across CPs. Though it is difficult to characterize confusion heterogeneity across HI listeners due to low row sums, threshold variability can be characterized for each listener even with the relatively low row sums. This is because the threshold variability is quantified using the saturation threshold SNR_{90} (i.e., the SNR at which the recognition score is 90%) and the confusion threshold SNR_g (i.e., the SNR at which a confusion group is formed). At SNR_{90}, there are very few confusions, and thus the probability distribution of the row is unimodal, which can be reliably estimated even with low row sums.

The saturation thresholds (SNR_{90}) are shown by vertical dash-dotted lines in Fig. 9. Listener 216L is unable to recognize /p/ (top right panel), resulting in almost 100% error for /p/ at all SNRs. In this case, 216L is considered to have a $SNR_{90}=\infty$ for consonant /p/. On the other hand, 208L's recognition score for /z/ (bottom right panel) is always greater than 90% at all tested SNRs. This is represented with a $SNR_{90}=-\infty$.

Figure 10 shows all SNR_{90} values for ANH (filled squares) and HI (open symbols) listeners for each consonant. The numbers at the top and bottom indicate the number of listeners with $SNR_{90}=\infty$ and $-\infty$, respectively, for each consonant. The SNR_{90} values falling between +12 dB SNR and the quiet condition are estimated by assuming a SNR of +18 dB for the quiet condition. Hence the ordinate scale between the two horizontal dashed lines in Fig. 10 is warped.

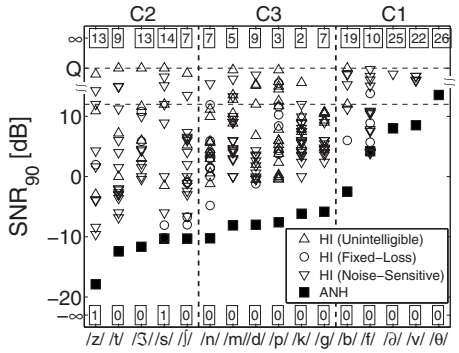


FIG. 10. Saturation thresholds SNR_{90} for ANH (filled circles) and HI (open symbols) listeners sorted by ANH SNR_{90} values. The numbers at the top and the bottom indicate the number of listeners with $SNR_{90}=\infty$ ($P_c<90\%$ at all SNR) and $SNR_{90}=-\infty$ ($P_c>90\%$ at all SNR) for each consonant. The shapes of open symbols represent the consonant loss categories derived from the CLPs. The categorization of consonants in sets C1, C2, and C3 is according to Phatak and Allen (2007).

As discussed earlier, a single performance point, such as SNR_{90} or SNR_{50}, cannot characterize the shape of consonant loss curves. Therefore, different symbols are used to indicate the CLP groups, which relate to the shapes of consonant loss curves. The symbol indicates whether a consonant was an unintelligible (△), a noise-sensitive (▽), or a fixed-loss (○) consonant for each listener.

Phatak *et al.* (2008) collected data on NH listeners and used the saturation threshold SNR_{90} as a quantitative measure of noise-robustness (Régnier and Allen, 2008). However, for HI ears, it is a measure of overall consonant loss, which includes both audibility loss and noise-sensitivity. The special case of $SNR_{90}=\infty$ (i.e., $P_c<90\%$ in quiet) indicates audibility loss. Thus the SNR_{90} values, combined with the CLP groups (given by the symbols), can provide both the amount and the nature of consonant loss for each listener on a consonant-by-consonant basis.

The consonants in Fig. 10 are divided into three sets that, with the exception of /m/, are same as the high-error set C1, low-error set C2, and average-error set C3 from Phatak and Allen (2007). Consonants from set C1 $=\{/b/,/f/,/ð/,/v/,/θ/\}$ are difficult to recognize for both ANH and HI listeners. The ANH SNR_{90} values for the entire set C1 are all greater than −3 dB, and on an average, more than 20 HI listeners have C1 $SNR_{90}=\infty$. On the other hand, ANH scores for set C2$=\{/z/,/t/,/ʒ/,/s/,/ʃ/\}$ are very high ($SNR_{90}<-10$ dB), but HI listeners have poor performance for these consonants too. On an average, more than 11 HI listeners have $SNR_{90}=\infty$ for set C2 consonants.

Table III shows that the mean PTA_{HF} values for HI listeners with $SNR_{90}=\infty$ are relatively higher for C2 consonants (mean $PTA_{HF}>60$ dB HL) than for the other two consonant sets. Phatak and Allen (2007) showed that the spectral energy in C2 consonants is concentrated at frequencies above 3 kHz, which could be obscured when the high-frequency audiometric thresholds are elevated. NH listeners benefit from this high-frequency speech information, which is not masked by a speech-weighted noise masker, but HI listeners are deprived of it due to their high-frequency hearing loss.

TABLE III. The number of HI listeners having $SNR_{90}=\infty$ and their mean PTA_{HF} value for each consonant.

Set	Consonant	HI group A (N=16)	HI group M (N=3)	HI group D (N=7)	Mean PTA_{HF} (dB HL)
C2	/t/	7	1	1	67.78
	/z/	12	0	1	66.35
	/ʒ/	12	1	0	64.62
	/s/	11	1	2	64.46
	/ʃ/	7	0	0	61.43
C3	/d/	7	1	1	63.06
	/n/	6	1	0	56.43
	/m/	5	0	0	55.50
	/k/	2	0	0	52.50
	/g/	6	1	0	51.79
	/p/	3	0	0	46.67
C1	/f/	8	1	1	59.00
	/v/	16	2	4	57.95
	/b/	13	2	4	56.71
	/ð/	16	2	7	57.90
	/θ/	16	3	7	57.50

IV. DISCUSSION

In this study we have analyzed the SNR differences between consonant recognition performances of HI and average NH listeners. Such an analysis was not possible with some of the past CM data on HI listeners, which were collected either in absence of masking noise (Walden and Montgomery, 1975; Bilger and Wang, 1976; Doyle *et al.*, 1981) or at a single SNR (Dubno *et al.*, 1982). The test-retest analysis shows that the results are consistent ($r>0.95$) for the six re-tested listeners (Fig. 1), indicating that the testing procedure is repeatable.

Festen and Plomp (1990) reported parallel *psychometric functions* of average NH (slope=21%/dB) and average HI (slope=20.4%/dB) listeners with sentences that suggest a score-independent constant SNR-loss. Data of Wilson *et al.* (2007) show that in multitalker babble masker, the relative slopes of NH and HI psychometric functions depend on the speech material. We found that the performance-SNR functions [i.e., P_e(SNR), Fig. 3(a)] of individual HI listeners for consonant recognition, measured using nonsense syllables in steady-state noise, are not parallel to the corresponding ANH function. Thus, the average consonant loss for individual HI listener is a function of the score [Fig. 3(b)]. This variation in SNR difference (i.e., the average consonant loss) is ignored in the traditional SRT-based measure of SNR-loss. The variation in consonant loss with performance level, indicated by the change in the slope of consonant loss curves, characterizes the nature of consonant loss (i.e., unintelligibility in quiet vs noise-sensitivity).

The SNR difference in HI and NH performances is consonant-dependent. Both HI and NH listeners have significant variability across consonants, which is obscured in an average-error measure by design. High consonant error does not necessarily imply high SNR difference for that consonant. A comparison with the corresponding ANH listener er-

ror is necessary for measuring the consonant-specific SNR deficit.[3] The CLP is a compact graphical comparison of each HI listener to the average NH listener on a consonant-by-consonant basis (Fig. 5). The slopes of CLP curves separate the unintelligible consonants from the noise-sensitive ones. The two types of consonant loss, i.e., unintelligibility and noise-sensitivity, are related to Plomp's (1978) *A*-factor (i.e., audibility) and *D*-factor (i.e., distortion or SNR-loss) losses, respectively. The CLPs reveal that the consonant loss for 16 of the 26 tested ears was dominated by the audibility loss, resulting in type A CLPs, which are characterized by average consonant loss curves with shallow slopes due to a majority of unintelligible consonants. Of the remaining ten HI ears, seven showed type D CLPs, which suggests that the consonant loss for these listeners was mostly due to noise-sensitivity or the SNR-loss. Listeners with type A CLP did not overcome audiometric loss as much as those with type D CLPs at their MCLs. However, there was no statistically significant difference between the audiograms of listeners with the two CLP types. A bilateral HI listener can have few but significant differences between CLPs for the two ears (Fig. 8). This suggests that the consonant loss depends, to some extent, on the differences in peripheral hearing other than the audiometric thresholds.

The audibility loss in HI listeners, quantified by large values of SNR_{90} (in many cases, $SNR_{90}=\infty$), was mostly restricted to the high-frequency C2 consonants (Fig. 10). While both ANH and HI listeners struggle with C1 consonants, only HI listeners with poorer high-frequency hearing performed poorly ($SNR_{90}=\infty$) for the high-frequency C2 consonants (Table III), which are high-scoring consonants for the NH listeners (Phatak and Allen, 2007). Other than this relationship between high-frequency loss and C2 consonants, the exact distribution of consonants in the three CLP categories does not correlate with, and therefore cannot be predicted from, the audiograms.

CLPs could be used in clinical audiology to obtain a simple snapshot of the patient's consonant perception. Collecting reliable confusion data is time consuming. The current CM experiment required approximately 3 h per subject. CLPs require only recognition scores (i.e., the CM diagonal element) and not individual confusions (i.e., the off-diagonal elements of CM). Reliable estimation of the recognition score does not require row sums as high as those required for estimating individual confusions. Also, CLP curves show significant variation in slope over a limited SNR range for HI listeners (i.e., for $SNR_{HI}>0$ dB). By reducing the number of tokens of each consonant and the number of SNRs, the testing time can be reduced to 5–10 min.

The information obtained from such a short clinical test can be used to customize the rehabilitation therapy for a hearing aid or a cochlear implant patient. HI listeners trained on lexically difficult words show a significant improvement in speech recognition performance in noise (Burk and Humes, 2008). The duration of such training is generally long, spanning several weeks. This long-term training could be made shorter and more efficient by increasing the proportion of words containing unintelligible and noise-sensitive consonants for each listener. The information from the CLPs might also be used to optimize signal processing techniques in hearing aids. For example, if the unintelligible or the noise-sensitive consonants share some common acoustic features, then the hearing aid algorithms could be modified to enhance those particular types of features over others. Such an ear-specific processing could further "personalize" each hearing aid fitting.

The SNR-loss is intended to characterize the suprathreshold level loss (i.e., Plomp's *D*-factor) and therefore one would expect it to be independent of the audio-metric thresholds. The flattening of average consonant loss curves at higher SNRs, due to the unintelligible consonants, indicates that audibility is not overcome at MCLs. It is impossible to know this information from a single performance point measure, such as ΔSNR_{50}. It also raises doubts about the clinically measured ΔSNR_{50}, using the QuickSIN recommended procedure, being a true measure of the audibility-independent SNR-loss. It is possible that even at levels as high as 85 dB SPL, speech may be below threshold at certain frequencies (Humes, 2007). Speech information is believed to have multiple cues, which may be redundant for a NH listener in the quiet condition (Jenstad, 2006). This redundancy would allow the HI listener to recognize speech even if elevated thresholds mask some speech cues. Such presentation levels may be incorrectly considered as suprathreshold presentation levels.

Another implication of this hypothesis is that the SNR-loss or the noise-sensitivity may be partially due to masking of noise-robust speech cues by elevated thresholds. Each speech cue has a different perceptual importance and noise-robustness. If the cue masked by the elevated thresholds is the most noise-robust cue for a given consonant, then the noise-robustness of that consonant will be reduced, turning it into a noise-sensitive consonant. For example, Blumstein and Stevens (1980) argued, using synthesized speech stimuli, that the release burst is the necessary and sufficient cue for the perception of stop plosives, and the formant movements are secondary cues. Using natural speech, Régnier and Allen (2008) clearly demonstrated that the release burst is also the most noise-robust cue for recognizing plosive /t/. At MCL, HI listeners with high-frequency hearing loss have high scores for consonants /p/, /t/, and /k/ in quiet, but not in the presence of noise. Under noisy conditions, /t/-/p/-/k/ is the most common confusion for these consonants, which is caused by the high-frequency hearing loss (see Sec. III E). Thus the elevated thresholds likely contribute to the noise-sensitivity. Therefore, it is critical to verify that the audiometric loss has been compensated with spectral shaping before estimating the SNR-loss.

Some past studies have concluded that the suprathreshold SNR-loss may be non-existent compared to the audibility loss (Lee and Humes, 1993; Zurek and Delhorne, 1987). Lee and Humes (1993) used meaningful sentences to measure the SNR-loss, which is the most common stimuli for measuring SNR-loss (Plomp, 1986; Killion *et al.*, 2004). Context, due to meaning, grammar, prosody, etc., can partially compensate hearing deficits. Therefore words and sentences, though easier to recognize than nonsense syllables, can underestimate the SNR-loss. The ΔSNR_{50} measure with

sentences are on an average 12 dB smaller than those measured with consonants (Fig. 2). Also, SNR-losses that exist for individual consonants, and which may exist for vowels, cannot be determined with an average measure like the word recognition score. Zurek and Delhorne (1987) used CV syllables, but they compared noise-masked normals with HI listeners. However, noise-masked elevated thresholds are not equivalent to a hearing loss, and additional hearing deficits exist in HI listeners that can affect speech perception (Humes *et al.*, 1986). In many cases, a HI listener performs better than the corresponding noise-masked NH listener. A more relevant and perhaps more accurate comparison would be between an HI listener with spectral gain compensation for hearing loss and an average NH listener. Humes (2007) showed that the performance deficits in noise exist even after carefully compensating the audiometric losses of HI listeners. He attributed this SNR-loss to aging and differences in cognitive and central processing abilities. However, this loss could also be due to poor spectral and temporal resolution in the peripheral auditory system. A loss of resolution would lead to degradation of the spectral and temporal cues, which could affect the performance of subsequent auditory processing tasks that are involved in speech recognition, such as integration of speech cues across time and frequency (Allen and Li, 2009).

Since the purpose of our experiment was to analyze the effect of hearing loss on consonant perception, no spectral correction was provided to compensate for the listener's hearing loss. Providing such correction will help the unintelligible consonants, but may not help the noise-sensitive consonants. Furthermore, the unintelligible consonants, after spectral correction, may not become low-loss consonants. Instead, they may still have noise-sensitivity due to hearing deficits other than the elevated thresholds. To answer these questions, this study is currently being repeated with spectrally corrected stimuli to compensate individual HI listener's hearing loss (Li and Allen, 2009).

V. CONCLUSIONS

The key results in this study can be summarized as follows.

(1) The CLP is a compact representation of the consonant-specific SNR differences over a range of performance for individual listeners (Fig. 5). It shows that the hearing loss renders some consonants unintelligible and reduces noise-robustness of other consonants. Audiometric loss affects some consonants at a presentation level as high as 85 dB SPL.
(2) The SNR difference between HI and NH performances varies with the performance level (i.e. recognition score) (Fig. 3). This variation is ignored in the traditional ΔSNR_{50} metric for SNR-loss, which is measured at a single performance point (i.e., $P_c=50\%$).
(3) The loss in consonant recognition performance is consonant-specific. HI listeners with poorer hearing at and above 4 kHz show more loss for high-frequency consonants /s/, /ʃ/, /t/, /z/, and /ʒ/ (Table III).
(4) Individual consonant losses, which determine the distribution of consonants in the CLP, cannot be predicted from the audiometric thresholds alone (Fig. 7).
(5) Consonant confusions of HI listeners vary across the listeners and are, in many cases, significantly different from the ANH confusions (Fig. 9). Some HI confusions are a result of the high-frequency hearing loss (Sec. III E).
(6) Sentences are not the best stimuli for measuring the SNR-loss as the context information in sentences partially compensates the hearing loss. The ΔSNR_{50} values are about 12 dB greater for consonants than for sentences (Fig. 2).

ACKNOWLEDGMENTS

This research was supported by a University of Illinois grant. Partial salary support was provided by NIDCD (No. R03-DC06810) and by Cooperative Research and Development Agreements between the Clinical Investigation Regulatory Office, U.S. Army Medical Department and School and the Oticon Foundation, Copenhagen, Denmark. The opinions and assertions presented are the private views of the authors and are not to be construed as official or as necessarily reflecting the views of the Department of the Army, or the Department of Defense. The data collection expenses were covered through research funds provided by Etymotic Research and Phonak. Portions of this work were presented at the IHCON Meeting (Lake Tahoe, CA) in 2006, at the ARO Midwinter Meeting (Denver, CO), the AAS Conference (Scottsdale, AZ), and the Aging and Speech Communication Conference (Bloomington, IN) in 2007, and in the Department of Hearing and Speech Science Winter Workshop at the University of Maryland (College Park, MD) in 2008. We thank Ken Grant and Harvey Abrams for informative discussions and insights. Mary Cord and Matthew Makashay helped in QuickSIN level measurements.

[1]With the recommended calibration procedures for QuickSIN (i.e., 0 VU input to the audiometer and 70 dB HL attenuator setting), the output level of the 1 kHz calibration tone is 88 dB SPL. At this level the most frequent peaks of the sentence stimuli are between 80 and 85 dB SPL. This measurement was done using GSI 61 digital audiometer, EARTONE 3A insert earphones, Zwislocki DB-100 coupler, and Larson Davis 800B sound level meter.
[2]Complete documentation is available at http://www.ldc.upenn.edu/Catalog/docs/LDC2005S22/doc.txt.
[3]While there is a variability in the consonant perception across the NH listeners, it is much greater across HI listeners. For example, the low-error, average-error, and high-error consonant sets, observed in study of Phatak *et al.* (2008), were the same for all NH listeners tested. Such an error-based consonant categorization is different for each HI listener. Therefore, the consonant CMs can be averaged across NH listeners to obtain ANH CMs, but such an average across HI listeners is statistically unstable.

Allen, J. B., and Li, F. (**2009**). "Speech perception and cochlear signal processing," IEEE Signal Process. Mag. **26**, 73–77.
Bilger, R. C., and Wang, M. D. (**1976**). "Consonant confusions in patients with sensoryneural hearing loss," J. Speech Hear. Res. **19**, 718–749.
Blumstein, S. E., and Stevens, K. N. (**1980**). "Perceptual invariance and onset spectra for stop consonants in different vowel environments," J. Acoust. Soc. Am. **67**, 648–662.
Boothroyd, A., and Nittrouer, S. (**1988**). "Mathematical treatment of context effects in phoneme and word recognition," J. Acoust. Soc. Am. **84**, 101–

114.
Burk, M. H., and Humes, L. E. (**2008**). "Effects of long-term training on aided speech-recognition performance in noise in older adults," J. Speech Lang. Hear. Res. **51**, 759–771.
Cooper, K. N., Delattre, P. C., Liberman, A. M., Borst, J. M., and Gerstman, L. J. (**1952**). "Some experiments on perception of synthetic speech sounds," J. Acoust. Soc. Am. **24**, 597–606.
Doyle, K. J., Danhauer, J. L., and Edgerton, B. J. (**1981**). "Features from normal and sensorinueral listeners' nonsense syllable test errors," Ear Hear. **2**, 117–121.
Dubno, J. R., Dirks, D. D., and Langhofer, L. R. (**1982**). "Evaluation of hearing-impaired listeners using a nonsense-syllable test II. Syllable recognition and consonant confusion patterns," J. Speech Hear. Res. **25**, 141–148.
Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.
Fousek, P., Svojanovsky, P., Grezl, F., and Hermansky, H. (**2004**). "New nonsense syllables database—Analyses and preliminary ASR experiments," in Proceedings of the International Conference on Spoken Language Processing (ICSLP), pp. 2749–2752.
French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.
Gordon-Salant, S. (**1985**). "Phoneme feature perception in noise by normal-hearing and hearing-impaired subjects," J. Speech Hear. Res. **28**, 87–95.
Helfer, K. S. (**1994**). "Binaural cues and consonant perception in reverberation and noise," J. Speech Hear. Res. **37**, 429–438.
Hudgins, C. V., Hawkins, J. E., Jr., Karlin, J. E., and Stevens, S. S. (**1947**). "The development of recorded auditory tests for measuring hearing loss for speech," Laryngoscope **57**, 57–89.
Humes, L. E. (**2007**). "The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults," J. Am. Acad. Audiol **18**, 590–603.
Humes, L. E., Dirks, D. D., Bell, T. S., and Kincaid, G. E. (**1987**). "Recognition of nonsense syllables by hearing-impaired listeners and by noise-masked normal listeners," J. Acoust. Soc. Am. **81**, 765–773.
Jenstad, L. (**2006**). "Speech perception and older adults: Implications for amplification," Proceedings of the Hearing Care for Adults, Chap. 5, pp. 57–70.
Killion, M. C. (**1997**). "SNR loss: I can hear what people say, I can't understand them," Hear. Rev. **4**, 8–14.
Killion, M. C., Niquette, P., Gudmundsen, G. I., Revit, L. J., and Banerjee, S. (**2004**). "Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **116**, 2395–2405.
Lee, L. W., and Humes, L. E. (**1993**). "Evaluating a speech-reception threshold model for hearing-impaired listeners," J. Acoust. Soc. Am. **93**, 2879–2885.
Li, F., and Allen, J. B. (**2009**). "Consonant identification for hearing impaired listeners," J. Acoust. Soc. Am. **125**, 2534.
Licklider, J. C. R. (**1948**). "The influence of internal phase relations upon the masking of speech by white noise," J. Acoust. Soc. Am. **20**, 150–159.
Lobdell, B., and Allen, J. B. (**2007**). "Modeling and using the vu-meter (volume unit meter) with comparisons to root-mean-square speech levels," J. Acoust. Soc. Am. **121**, 279–285.
Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.
Phatak, S. A., and Allen, J. B. (**2007**). "Consonant and vowel confusions in speech-weighted noise," J. Acoust. Soc. Am. **121**, 2312–2326.
Phatak, S. A., Lovitt, A., and Allen, J. B. (**2008**). "Consonant confusions in white noise," J. Acoust. Soc. Am. **124**, 1220–1233.
Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (**1995**). "How young and old adults listen to and remember speech in noise," J. Acoust. Soc. Am. **97**, 593–608.
Plomp, R. (**1978**). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am. **63**, 533–549.
Plomp, R. (**1986**). "A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired," J. Speech Hear. Res. **29**, 146–154.
QuickSIN Manual (**2006**). QuickSIN Speech-in-Noise Test (Version 1.3), Etymotic Research, Inc., 1.3 ed.
Régnier, M., and Allen, J. B. (**2008**). "A method to identify noise-robust perceptual features: Application for consonant /t/," J. Acoust. Soc. Am. **123**, 2801–2814.
Tillman, T. W., and Olsen, W. O. (**1973**). "Speech audiometry," in *Modern Developments in Audiology*, 2nd ed., edited by J. Jerger (Academic, New York), pp. 37–74.
Walden, B. E., and Montgomery, A. A. (**1975**). "Dimensions of consonant perception in normal and hearing-impaired listeners," J. Speech Hear. Res. **18**, 444–455.
Wilson, R. H., McArdle, R. A., and Smith, S. L. (**2007**). "An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss," J. Speech Lang. Hear. Res. **50**, 844–856.
Zurek, P. M., and Delhorne, L. A. (**1987**). "Consonant reception in noise by listeners with mild and moderate sensorineural hearing loss," J. Acoust. Soc. Am. **82**, 1548–1559.

Extraction of bowing parameters from violin performance combining motion capture and sensors

E. Schoonderwaldt[a)]
Department of Speech Music and Hearing, School of Computer Science and Communication, KTH, Lindstedtsvägen 24, SE-100 44 Stockholm, Sweden and Input Devices and Music Interaction Laboratory, Schulich School of Music, McGill University, 555 Sherbrooke Street West, Montreal, Quebec H3A 1E3, Canada

M. Demoucron
IRCAM, Centre Pompidou, CNRS UMR 9912, 1 Place Igor Stravinsky, 75004 Paris, France

(Received 19 June 2008; revised 1 July 2009; accepted 23 August 2009)

A method is described for measurement of a complete set of bowing parameters in violin performance. Optical motion capture was combined with sensors for accurate measurement of the main bowing parameters (bow position, bow velocity, bow acceleration, bow-bridge distance, and bow force) as well as secondary control parameters (skewness, inclination, and tilt of the bow). In addition, other performance features (moments of on/off in bow-string contact, string played, and bowing direction) were extracted. Detailed descriptions of the calculations of the bowing parameters, features, and calibrations are given. The described system is capable of measuring all bowing parameters without disturbing the player, allowing for detailed studies of musically relevant aspects of bow control and coordination of bowing parameters in bowed-string instrument performance. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3227640]

PACS number(s): 43.75.De, 43.75.Yy [NHF] Pages: 2695–2708

I. INTRODUCTION

A. Background

The interaction between the performer and the instrument is a fascinating subject of study. Technological developments have successively opened up new ways of capturing this interaction, shedding more light on tone control and musical performance aspects. An important example from the last two decades is the computer-controlled grand piano, such as Yamaha's Disklavier or the Bösendorfer SE290, which offers built-in technology for capturing key played, hammer timing, and key velocity, the main parameters for studying piano performance.[1]

In violin performance, the sound originates from the frictional interaction between the bow and the string, and the player exerts continuous and minute control of the tone, basically by varying bow velocity, bow force, and bow-bridge distance. The production of a musical tone and mastery of a wide range of bowing techniques requires a high degree of skill by the player, acquired during many years of training and practicing. The fact that the tone is mainly controlled by bowing action has the advantage that most control parameters are directly accessible for measurements, making bowed-string instrument performance highly suitable for detailed study of performer-instrument interaction.

The first detailed recordings of bowing gestures were made in the 1930s by Hodgson,[2] who visualized the spatial trajectories of the bow and the player's bow arm in cyclegraphs (a photographic record of the track covered by a moving object). His findings led to new, sometimes controversial, insights[3] in the way bow movements were executed in practice with important pedagogical implications.

The first setup for calibrated measurement of the main bowing parameters (bow position, velocity, force, and bow-bridge distance in violin playing) was developed by Askenfelt.[4,5] A thin resistive wire inserted among the bow hairs allowed measurement of both transverse bow position and bow-bridge distance by means of two Wheatstone bridges, connected at the point of bow-string contact. Bow force was measured using strain gauges at both ends of the bow-hair ribbon. The setup was used to study typical bowing patterns in musical performance as well as strategies in playing different dynamics and expressive renderings.

More recently, sensor-based systems for capturing bowing gestures have been developed, often for use in electronic interactive performances, for example, the Hypercello,[6] the BoSSA combined with the R-bow,[7] the vBow,[8] the hyperbow[9–11] (further developed from the Hypercello), the overtone violin,[12] and the augmented violin.[13] This type of extended instruments or controllers generally includes sensors for measurement of bow position, bow acceleration, and/or bow force, and the signals based on the performer's actions are used to control sound synthesis or other effects.[6,14] The focus of these systems is mostly on reliable feature extraction in live performance, less on accuracy and absolute calibration.

The hyperbow has also been used as a research tool for analysis of bowing gestures.[15] In its latest stadium of development the hyperbow features calibrated sensors for measur-

[a)]Author to whom correspondence should be addressed. Electronic mail: schoondw@kth.se

ing the position and orientation of the bow relative to the violin, along with bow acceleration and bow force. Bow position and bow-bridge distance are measured by electric field sensing.[16] Bow force is measured in two directions, downward (normal to the string) and lateral (orthogonal to the string), using two pairs of strain gauges mounted at the middle of the bow stick. The orientation of the bow and the violin is determined using six degree-of-freedom (6DOF) inertial measurement units (IMUs) consisting of gyroscopes measuring three directions of angular velocity and a three-axis accelerometer. A classification study using principal component analysis and machine learning techniques showed that the data generated by this system could be used to distinguish between different bowing patterns (détaché, martelé, and spiccato) and players.[15] The setup has also been used for testing the playability of a violin physical model in real time as well as using recorded performance data.[15,17] In the recorded performances the hyperbow had to be complemented with an optical motion capture system for accurate measurement of bow position and bow velocity.

The augmented violin is similar in concept as the hyperbow, but simpler, not including gyroscopes and built-in sensors for measuring bow force.[13] Also the augmented violin has been used to study bowing patterns, focusing on features in the bow acceleration signal.[18,19] More recently, a "light" version of the augmented bow (accelerometers only) was used in combination with an optical motion capture system for measurement of bowing parameters (bow position, velocity, and acceleration) including the movements of the bow arm to study organizational aspects of motor control as function of tempo in détaché bowing.[20]

Other studies of bowed-string instrument performance have been performed using three dimensional (3D) motion capture techniques for the analysis of timing and coordination[21,22] and kinematics and kinetics related to the development of vocational injuries in musicians.[23–25] Further applications of motion capture, often in combination with visualization, have been developed for educational purposes.[26–28]

A promising alternative to optical motion capture is electromagnetic tracking. A single sensor provides both position and orientation (6DOF) so that only two sensors are needed for a complete spatial determination of the bow and the violin. This technique has been used by Peiper *et al.*[26] for classification of bowing patterns (referred to as articulations) and visualization of bowing gestures in a virtual reality environment. More recently, Maestre *et al.*[29] and Perez *et al.*[30,31] used an electromagnetic tracking device for measuring bowing parameters in violin performance for application in gesture-based synthesis and sound transformations. They were also able to obtain an estimation of bow force from the motion capture data without an additional sensor.

B. Aims

The method described in this paper was developed for detailed analysis of the relationship between physical aspects of bow-string interaction and the actions of the player in real violin performance, including the production of "steady" tones, as well as transients during attacks and bow changes. An additional goal was to shed more light on the player's strategies to shape the performance, more specifically, how the player manages to control *changes*, for example, in dynamics, in a musical context. The bow angles tilt, inclination, and skewness seem to play an important role therein. These angles are used for instantaneous control of details of the note played, like the effective width of the bow hair, and which string is being played. On a higher level and a longer time scale they play a major role in the pre-planning of the upcoming bowing gestures, like bow changes, string crossings, and changes in dynamic level.

A primary requirement on the method was therefore an accurate and precise acquisition of the main bowing parameters: bow velocity, bow-bridge distance, bow force, and acceleration. Furthermore, an exact determination of the orientation of the violin and the bow was required for calculation of the bow angles. For a reliable and accurate measurement of all these parameters it was decided to use a combination of optical motion capture to determine the position and orientation (pose) of the bow and the violin, and sensors on the bow for measuring bow force and acceleration.

Important additional requirements were that any regular bow and instrument should be possible to use, that the mass added to the bow was kept to a minimum, and that the measurements did not interfere with normal playing conditions. Optical motion capture offers a further possibility to include the body movements of the player, allowing to extend the study to instrumental and ancillary gestures.

After a short description of the experimental setup in Sec. II, the measurement of the pose of the violin and the bow will be discussed in Sec. III. This was done using the kinematic modeling and fit facilities in the Vicon software. The models were designed to allow straightforward and accurate calculation of the spatially defined bowing parameters (bow position, velocity, and bow-bridge distance) and bow angles (tilt, inclination, and skewness). Further, it will be shown how the models were extended for the determination of other features, such as bowing direction (down-bow and up-bow), string contact, and string played. In Sec. IV it will be explained how motion capture and sensor data were combined to obtain calibrated measurements of bow force and acceleration and remove the gravity component from the acceleration signal. Finally, the measurements will be illustrated with an example, and a comparison—as far as possible—will be made with some of the existing systems mentioned above. A large study of bow control in violin and viola performance with the use of the described system is reported in a companion paper.[32]

II. EXPERIMENTAL SETUP

A schematic overview of the used setup is shown in Fig. 1. Motion capture data were recorded via a Vicon data station connected to a personal computer (PC) running the Vicon acquisition software (Vicon workstation), and sensor data were recorded via a National Instruments data acquisition card on a second PC. The sound was synchronously recorded with the motion capture data as analog data via the

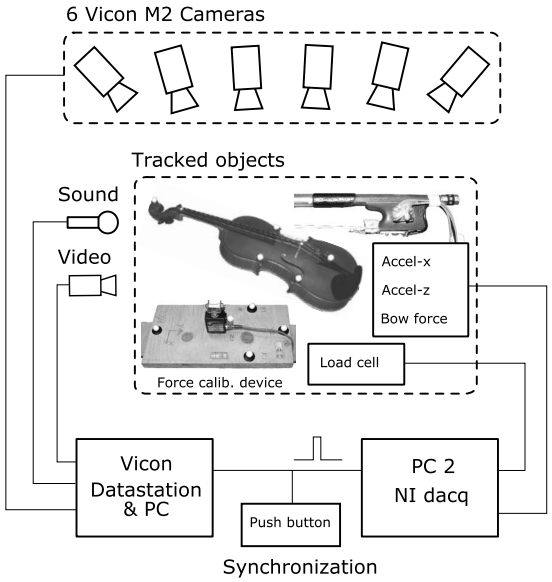


FIG. 1. Scheme of the setup used for measuring bowing gestures. Motion capture data and sensor data were recorded by separate data acquisition units. Synchronization of the data was achieved *a posteriori* via the signal from a push button. The included pictures of the tracked objects show the bow equipped with bow-force sensor and accelerometer, the violin (no additional sensors), and the device used for calibration of the bow-force sensor equipped with a load cell.

built-in acquisition card of the Vicon data station at a sample frequency of 40 kHz. Synchronization of the motion capture and sensor data was achieved with a push button providing a pulse at the beginning and the end of each recording (typically 1 min of duration). The push button signal was recorded on additional channels on both data acquisition units, allowing for *a posteriori* synchronization of the data with an accuracy of about 1 ms.

The sessions were also filmed with a digital camcorder directly connected to the Vicon PC by firewire. Acquisition of video was started and stopped in synchrony with the motion capture data.

A. Motion capture

A Vicon 460 system with six M2 cameras (maximum resolution of 1280×1024 pixels) was used for optical motion capture.[33] The cameras were positioned around the subject at distances of about 3–4 m. The frame rate used was 250 Hz.

It should be noted that the bow and the violin could be tracked with a better resolution with shorter camera distances. However, the used setup allows for full body motion capture of the player as well, the achieved resolution still being acceptable, as will be shown in Sec. III F.

Middle-sized reflective spherical markers (diameter of 10 mm and mass of 0.8 g) were attached to the bow and the violin using a special adhesive putty (Schertler Audio), leaving no traces on the varnish. Markers were positioned so that the violin sound would not be influenced (i.e., not on the bridge and on sound-radiating surfaces of the instrument). Additional markers used for calibration were small hemispherical markers of 3 mm diameter, which could be positioned accurately (within about 0.5 mm) on specific locations on the violin and the bow.[34] Further details of the marker placement and kinematic models of the violin and the bow will be discussed in Sec. III A.

B. Bow force and acceleration

Sensors were mounted on the bow for accurate measurement of bow force and acceleration. A three-axis accelerometer (STMicroelectronics LIS3L02AS4) with a linear measurement range of $\pm 60\ \mathrm{m/s^2}$ and frequency range down to dc was used for measuring bow acceleration in the longitudinal and vertical directions ($\hat{\mathbf{x}}_b$ and $\hat{\mathbf{z}}_b$ in Fig. 3). The wide measurement range was needed for reliable acquisition of bow acceleration. In a pilot study bow accelerations up to about $60\ \mathrm{m/s^2}$ were observed in *forte* 16th-note passages.

Bow force was measured using a custom-made sensor which registered the deflection of the bow hair at the frog. The sensor consisted of a leaf spring with strain gauges on both sides. The sensor was mounted rigidly to the frog and connected to the bow hair via a cylindrical bearing piece. A more detailed description of the principles and development of the bow-force sensor is provided by Demoucron.[35] To prevent possible damage to the violin due to the sensor under the frog, the violin was protected with a plastic C bout protector.

The bow-force sensor was calibrated using a miniature load cell with a capacity of about 25 N (Transducer Techniques MDB-5). The load cell in turn was calibrated using a set of standard weights (mass of 10–200 g).

The sensor signals were recorded at a sample rate of 10 kHz using a National Instruments acquisition card (NI PCI-4472B).

C. Playing comfort

The total mass added to the bow was estimated to be about 15 g (force sensor including mounting piece of 4 g, accelerometer of 3 g, "antennas" of 2.5 g (see Sec. III A), five reflective markers of 4 g, and putty of 1.5 g). Most of the weight (about 9 g) was added to the frog, where it influenced playing properties, such as moment of inertia relative to the hand of the player, the least. For comparison, the combined mass of the frog and the frog screw is about 17 g. The rest of the additional mass was placed in the lower half of the bow stick just below the balance point (about 5 g) and at the tip (about 1 g). The added components shifted the balance point about 2 cm toward the frog.

The wires from the sensors were taped to the player's arm to keep them out of the way and restrict their movement in dynamic playing situations. After a short period of familiarization all participating players in the recordings and pilot tests (about ten) confirmed that they felt comfortable in playing. One player commented on the stiffness of the wires, which he could notice during spiccato playing.

III. MOTION CAPTURE OF BOWING GESTURES

A. Kinematic models and marker configurations

The Vicon iQ software used for processing the motion capture data facilitates the use of kinematic models. A kine-

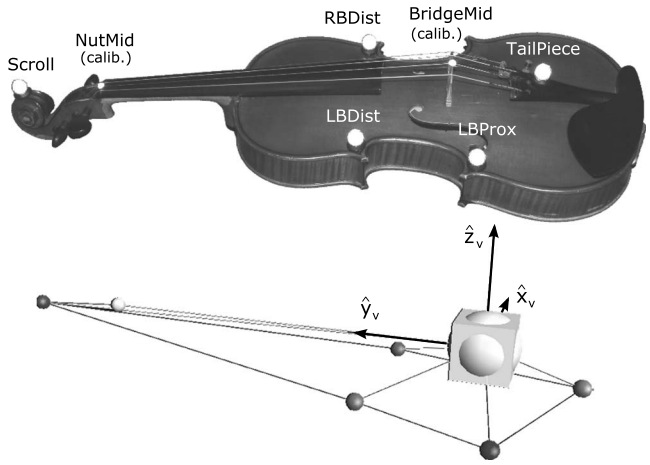


FIG. 2. Marker configuration for the violin and the corresponding kinematic model as implemented in the Vicon software. The five markers used for tracking the pose of the violin are labeled *Scroll*, *RBDist*, *LBDist*, *LBProx*, and *TailPiece*. The two calibration markers *BridgeMid* and *NutMid*, indicating the positions of the bridge and the nut between the D and A strings, are present only during calibration. The violin is modeled as one segment (rigid body), where the origin, indicated by the cube, coincides with *BridgeMid*. The orientation of the local reference frame is indicated by the basis vectors ($\hat{\mathbf{x}}_v$, $\hat{\mathbf{y}}_v$, and $\hat{\mathbf{z}}_v$).

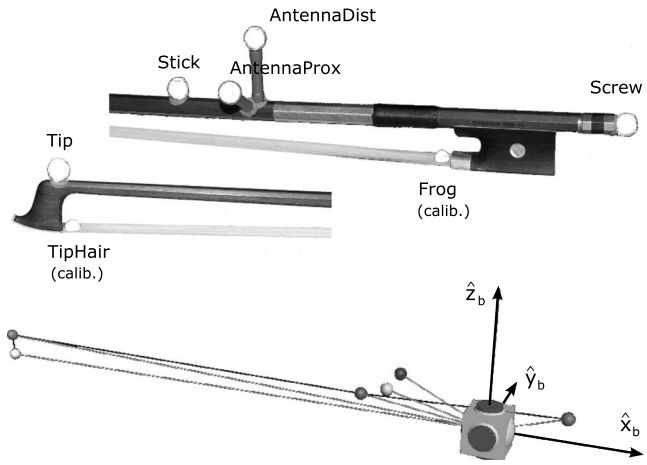


FIG. 3. Marker configuration for the bow and the corresponding kinematic model. The frog segment, indicated by the cube, is considered as a rigid body connected to four markers: *Stick*, *AntennaProx*, *AntennaDist*, and *Screw*. The 2DOF joint used for modeling bending of the stick coincides with the origin of the frog segment (defined by the *Frog* marker). The orientation of the stick segment relative to the frog is determined by the *Tip* marker. The two calibration markers *Frog* (frog segment) and *TipHair* (stick segment) indicate the positions of the hair terminations at the frog and tip. The orientation of the bow is indicated by the basis vectors ($\hat{\mathbf{x}}_b$, $\hat{\mathbf{y}}_b$, and $\hat{\mathbf{z}}_b$).

matic model typically consists of segments interconnected with joints with a certain number of DOFs. The segments are typically associated with one or more markers, defining the marker configuration of the kinematic model. Kinematic models can be built in the modeling module of the Vicon iQ software. The axes of the local reference frame of a modeled object can be fixed via geometrical constraints.

The marker configuration and kinematic model used for the violin are shown in Fig. 2. The violin is modeled as a rigid body (one segment). Five markers are used for tracking the pose of the violin. Two extra calibration markers are located at the middle of the bridge (*BridgeMid*) and at the nut (*NutMid*) between the D and A strings. In the kinematic model the origin corresponds to *BridgeMid*, and the *y*-axis corresponds to the line between *BridgeMid* and *NutMid* (virtual string). The *x*- and *z*-axes are defined by an additional constraint that the markers at the upper C bouts (*RBDist* and *LBDist*) share the same *z*-coordinate in the local reference system.

For the bow, a rigid body approximation is not appropriate as the stick can bend considerably under normal playing conditions.[35–37] Depending on tilt, bending can take place in two directions. Torsion of the stick can be neglected as the mechanical lever (height of the tip) is much smaller than that for bending (length of the stick).

The bow was modeled as compound of two (rigid) segments, the frog (root segment) and the stick, connected with a two DOF (2DOF) joint (Fig. 3). The joint coincided with the origin of the root segment (marker *Frog* at the ferrule), securing a fixed distance between the frog and tip (length of the bow hair). The frog segment included the lower, thick part of the stick from the frog screw (*Screw*) to some centimeters in front of the wrapping (*Stick*). In order to avoid a collinear configuration and to achieve a better spread between markers, two of the markers were placed on short antennas mounted on the stick. This configuration allowed for a complete measurement of the orientation of the bow, including bow tilt (rotation around the *x*-axis). The orientation of the stick segment relative to the frog segment was determined by a single marker (*Tip*). Two extra calibration markers were located on the bow-hair ribbon close to the frog (*Frog*) and the tip (*TipHair*). The calibration marker *Frog* defines the origin of the model, and the line between *Frog* and *TipHair* defines the *x*-axis. The *y*- and *z*-axes are established via the constraint that both *Stick* and *Screw* must lie in the *x*-*z* plane ($y=0$) in the local reference system.

The pose of the violin and the bow was determined by fitting the kinematic models to the measured 3D marker positions using the kinematic fit facilities in the Vicon iQ software and exported for further processing in MATLAB.[38] The orientations (Euler angles) were converted to rotation matrices.[39] The rows of these rotation matrices correspond to the basis vectors of the local reference systems expressed in world coordinates.[40] These basis vectors were extensively used in the calculations of the bowing parameters.

B. Calculation of bowing parameters

Having defined the local reference systems of the bow and the violin, the bowing parameters can be calculated based on these 6DOF representations, instead of the 3D positions of individual markers. In Fig. 4(a) the geometric relations used for the calculation of bowing parameters are shown. The bow-string contact point P is defined as the intersection between the line corresponding to the (virtual) string (direction $\hat{\mathbf{y}}_v$) and the bowing plane BPP'. The point P' is the projection of P on the bow.

Bow-bridge distance (the distance between the contact point and the bridge) is calculated as $\overrightarrow{VP}$, projected on the string direction, where V is the origin of the violin (marker *BridgeMid* at the top of the bridge, see Fig. 2):[41]

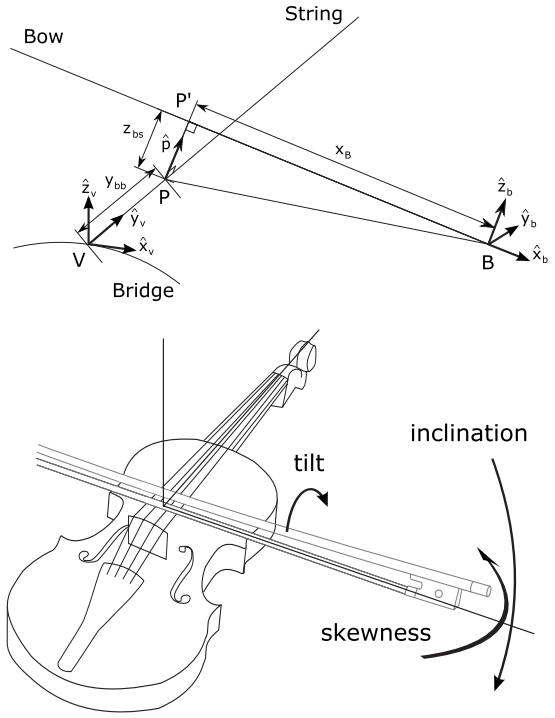


FIG. 4. (a) Geometric relations for calculation of bow-string contact point, bow position (x_B), bow-bridge distance (y_{bb}), and bow-string distance (z_{bs}). The points indicated are the local origins of the violin V (marker *BridgeMid* in Fig. 2) and the bow B (marker *Frog* in Fig. 3), the contact point on the string (P), and its projection on the bow (P'). Unit vector $\hat{\mathbf{p}}$ is perpendicular to both the bow and the string. Basis vectors of the local reference systems of the violin (subscript v) and the bow (subscript b) are indicated as well. (b) Bow angles skewness (φ), inclination (ϑ), and tilt (ψ). The arrows indicate the positive direction according to the definitions.

$$y_{\mathrm{bb}} = (P - V) \cdot \hat{\mathbf{y}}_v. \tag{1}$$

The bow-string distance (the height of the bow above the string) is defined as

$$z_{\mathrm{bs}} = (B - V) \cdot \hat{\mathbf{p}}, \tag{2}$$

where B is the origin of the bow (marker *Frog* at the ferrule, see Fig. 3) and $\hat{\mathbf{p}}$ is the unit vector perpendicular to both the bow and the string.

Bow position (the distance between the contact point and the frog) is defined as

$$x_B = (B - V) \cdot \hat{\mathbf{x}}_b, \tag{3}$$

which corresponds to the distance between P' and B. It should be noted that the calculated "distances" above can become negative, as they are projections on specific (non-static) axes.

Regarding the basis vectors of the bow the kinematic model used offers the choice between the frog and the stick segment. Typically, the basis vectors of the stick are preferred for the calculation of the bowing parameters because they correspond better to the line of the bow-hair ribbon when the stick is bent, and they can be measured more accurately (see Sec. III F). However, in some cases it might be worthwhile to include the bending of the stick in the evaluation, e.g., for estimation of bow force in the way proposed in Ref. 29 (see Sec. V). In such cases the basis vectors of the frog segment are more appropriate.

Bow velocity is defined as the velocity of the bow relative to the violin projected on the x-axis of the bow

$$v_B = \frac{\mathrm{d}(B - V)}{\mathrm{d}t} \cdot \hat{\mathbf{x}}_b. \tag{4}$$

Even if the bow velocity can be considered meaningful as a bowing parameter only when the bow is in contact with the string, the above definition of bow velocity is more general and allows for studying anticipatory movements of the bow in the air, like in bouncing bowing techniques such as spiccato.

For the numerical differentiation a central difference algorithm was used. The velocity signal can become rather noisy due to amplification-by-differentiation of the noise present in the measured positions. The quality of the velocity signals can be improved by low-pass filtering without losing essential details. It was found that a second-order Butterworth filter with a cut-off frequency of 18 Hz (applied back-and-forth to avoid phase shift) effectively filtered the noise, while preserving the shape of the velocity profile in fast bowing.

C. Calculation of bow angles

The three angles of the bow relative to the violin, *skewness* (φ), *inclination* (ϑ), and *tilt* (ψ), were defined in a way which reflect basic concepts in bowing, as shown in Fig. 4(b).[42]

Skewness is defined as the deviation of the bowing direction from orthogonality to the string. Considering turning the frog away from the player as the positive direction, it can be calculated as

$$\varphi = \frac{\pi}{2} - \arccos(\hat{\mathbf{y}}_v \cdot \hat{\mathbf{x}}_b). \tag{5}$$

Inclination is the angle associated with playing different strings. Adopting the convention that inclination increases from the lower to the higher-pitched strings it can be expressed as

$$\vartheta = \arccos(\hat{\mathbf{z}}_v \cdot \hat{\mathbf{x}}_b) - \frac{\pi}{2}. \tag{6}$$

Tilting is used to change the contact properties of the bow hair with the string by controlling the effective width of the bow hair in contact with the string and introducing a bow-force gradient across the width of the bow-hair ribbon. Here, tilt is defined as the angle between a plane parallel to the length axis of the bow ($\hat{\mathbf{x}}_b$) and the string ($\hat{\mathbf{y}}_v$), and a line with direction $\hat{\mathbf{y}}_b$. For convenience, the tilt direction used in classical playing with the stick leaning toward the fingerboard is taken as the positive direction. Tilt is zero when the hair is flat on the string. The expression for tilt then becomes

$$\psi = \arccos\left(\frac{\hat{\mathbf{x}}_b \times \hat{\mathbf{y}}_v}{|\hat{\mathbf{x}}_b \times \hat{\mathbf{y}}_v|} \cdot \hat{\mathbf{y}}_b\right) - \frac{\pi}{2}. \tag{7}$$

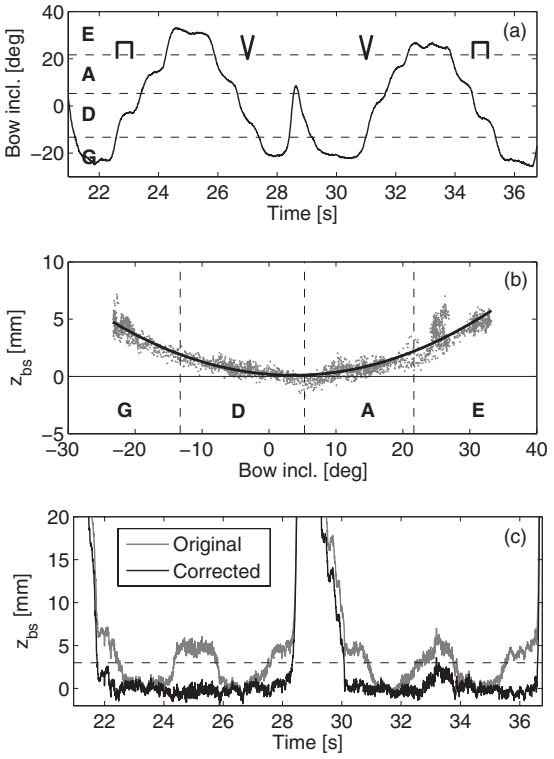


FIG. 5. Open-string arpeggios (played down-bow, up-bow, and reversed) used for the estimation of bow-string distance correction parameters. (a) Inclination of the bow versus time. The inclination ranges corresponding to the four strings are indicated by their pitches. (b) Bow-string distance as calculated by Eq. (2) versus bow inclination and fitted parabola. (c) Estimated bow-string distance (corrected and uncorrected). The dashed line indicates a typical threshold for the determination of bow-string contact.

D. Extending the instrument model

The geometrical model of the violin as described above features only one virtual string between the *BridgeMid* and *NutMid* markers. For a more complete definition of the violin model including four strings an additional calibration is needed. The string-crossing angles specific to an instrument can be obtained from a so-called *tuning trial*, in which the three double-string combinations (G-D, D-A, and A-E) are played slowly using the whole bow. The string-crossing angles can then be calculated by taking the average bow inclination (in violin coordinates) for each combination of strings. The standard deviations provide angular intervals, which can be useful for further feature extraction (see Sec. III E). When double strings are played softly (*pp*) the inclination of the bow is rather constrained in order to make both strings speak properly. However, the measured standard deviation will to a certain extent depend on the skill of the player.

A drawback of the one-string model is that the bow-string distance, as calculated by Eq. (2), deviates from the actual value depending on which string is played. The dependence of the bow-string distance on bow inclination can be determined via slow, whole bow arpeggios across all four strings (Fig. 5). It was found that the relation was well described by a second-order polynomial. The coefficients obtained by a least-squares fit can be used for improving the estimation of bow-string distance, as demonstrated in Fig. 5(c).

Also the other actual distances (bow position and bow-bridge distance) can deviate from those calculated using the one-string model, and corrections should be applied when an exact estimate is required. Regarding bow-bridge distance, a nonzero skewness of the bow results in a deviation of the calculated contact point along the string depending on which string is played. For example, at a skewness angle of 10° the deviation in bow-bridge distance amounts to about 2.5 mm for the outer strings. As skewness angles of this order can be observed in certain playing situations (see, e.g., Ref. 43) compensation might be needed, using knowledge of which string is played (see Sec. III E).

Given the specific string-crossing angles, the positions of the four strings relative to the local reference system of the violin can be determined when the distances between the adjacent strings are known. For most violins the separation between the strings at the bridge is about 11 mm. The calculated (or measured) string positions may be used as alternative origins for the calculation of bowing parameters, removing the need for additional corrections. However, a drawback of the use of multiple origins in the calculations is the presence of discrete jumps at the moments of string crossings, causing artifacts in, e.g., calculated bow velocity. Furthermore, the choice of the origin is dependent on the, sometimes uncertain, decision of which string is played. In the following the one-string model is therefore preferred, applying corrections afterward when necessary.

E. Feature extraction

With the use of the calculated bowing parameters and the bow-instrument model, extraction of performance features such as bowing direction, string contact, and string played becomes straightforward.

Bowing direction (up-bow and down-bow) is simply determined by the sign of the bow velocity. For detection of zero bow velocity a threshold is needed (about 1–1.5 cm/s with the current setup), taking the noise fluctuations into account. Bowing direction is set to zero when the bow is not in contact with the string.

To determine if the bow is in contact with the string a threshold criterion (3–5 mm above the string) is applied to the corrected bow-string distance [dashed horizontal line in Fig. 5(c)].

The string played is determined by bow inclination [Eq. (6)]. Also in this case a threshold method is used based on the specific string-crossing angles of the instrument. Detection of double-string playing is tricky as this involves angular regions of finite width around the string-crossing angles. The width of the transition region is a trade-off between sharpness in the detection of string crossings and robust identification of double-string passages. A useful way to define the width is based on the measured standard deviation of the string-crossing angles in the tuning trial. Preliminary tests indicated that at least a 99% confidence interval (obtained by multiplying the standard deviation by a factor of 2.58) was needed for satisfactory identification of double stops.

The application of threshold methods to more or less

TABLE I. Average noise rms and peak-to-peak values of the bow angles tilt, inclination, skewness, and distances between the bow and the violin.

Bowing parameter	rms	Peak-to-peak
Tilt (deg)	0.6	4
Inclination (deg)	0.2	1
Skewness (deg)	0.08	0.5
Bow pos. (mm)	0.4	2
Bow-bridge dist. (mm)	0.5	3
Bow-string dist. (mm)	0.3	2

noisy signals will result in classification errors. These can be effectively reduced by applying a running median filter with a time window in the order of the shortest expected note duration.

F. Precision and accuracy

1. Noise

Noise estimations were made from the slow four-string arpeggio recording. The rms noise of each coordinate was calculated after filtering the slowly varying signals using a high-pass filter (Butterworth, order of 4 and cut-off of 8 Hz). The noise distributions of the coordinates and angles were well described by a Gaussian distribution.

The rms noise of the marker positions was well below 0.5 mm, corresponding to peak-to-peak values of less than 3 mm. The rms noise in the positions of the origins of the fitted models was somewhat less (0.2–0.4 mm).

Table I indicates how the noise is propagated to the calculated bowing parameters. It can be seen that tilt is relatively noisy compared to the other bow angles. This is due to the marker configuration for the bow: as the mounted antennas are short compared to the length of the bow the rotation about the length axis is more sensitive to noise in marker positions. The noise rms of the relative distance measures remained well below 1 mm. For clarity the peak-to-peak noise is also displayed, obtained by multiplying the rms value by a factor of 6.18 (99.9% interval of the noise distribution).

In order to assess a possible influence of bow velocity on the accuracy of marker positions, the distance between the *Stick* and the *Screw* markers on the bow was analyzed in a highly dynamic trial with bow velocities of more than 2 m/s and accelerations of 60 m/s^2. No systematic relation was found between marker distance and bow speed. The total (unfiltered) noise rms of distance was 0.8 mm, similar to the slow arpeggios. These results indicate that marker positions could be measured with acceptable accuracy and precision even at extreme bow velocities and accelerations.

For comparison, the precision (rms distance between markers) reported for a similar motion capture system was about 15 μm under the most favorable circumstances. The large discrepancy can be explained by the suboptimal conditions under which motion capture was performed, including the circular camera configuration (an umbrella configuration is recommended for optimal results), the relatively large camera distances, and the laboratory environment (normal office, not optimized for motion capture). Furthermore, the reflections of the violin might have degraded the performance of the motion capture system.

The achieved noise levels were considered acceptable for the purpose of the current measurements. When needed, appropriate smoothing techniques were applied, for example, for the calculation of bow velocity.

2. Bow-bridge distance

To estimate the accuracy of bow-bridge distance due to the uncertainty in the placement of the calibration markers, static trials were performed using a test bench equipped with a ruler. The dimensions and marker configuration of the test bench were similar to those of the violin. The results showed that there was no systematic deviation of bow-bridge distance, and the accuracy error was well within 1 mm.

A varying error in measured bow-bridge distance during playing arises from bow tilt. When the bow is tilted, the center line of the part of the bow-hair ribbon, which is in actual contact with the string, will deviate from the bow-hair line in the model. The deviation is dependent on tilt angle and bow force. In an extreme case with a tilt angle of 45° and a very low bow force (one hair in contact with the string) the deviation can reach about 3–4 mm. However, under normal playing conditions the deviation will be smaller than 1 mm in most cases.

All the errors discussed above are small compared to the width of the bow-hair ribbon (8–10 mm), which forms a fundamental uncertainty in the determination of bow-bridge distance.

IV. MEASUREMENT OF ACCELERATION AND BOW FORCE

Sensors were attached to the bow for measuring bow acceleration and bow force. The data obtained via motion capture and sensors could be considered complementary. In particular, the signal-to-noise ratio of the second derivative of position data is usually rather poor and requires a fair amount of smoothing. The accelerometer signal contains much finer details and is therefore preferred in studies of transients, e.g., during changes in bowing direction (bow changes). On the other hand, the advantage of motion capture systems to measure position and orientation of objects accurately can be used for calibrating the accelerometer and other sensors.

A. Acceleration

Besides measuring the actual bow acceleration, the accelerometer was sensitive to the inclination of the bow due to the influence of gravity. As the inclination θ (in world coordinates) is measured by the motion capture system this information can be used to calibrate the accelerometer and remove the gravity component $[a_0(\theta) = -9.81 \sin \theta]$ from the accelerometer signal. The calibration measurement was performed by rotating the bow slowly around the $\hat{\mathbf{y}}_b$-axis, varying the inclination of the two accelerometer axes $\hat{\mathbf{x}}_b$ and $\hat{\mathbf{z}}_b$. The gain k and offset s_0 of the accelerometer signal s were obtained via

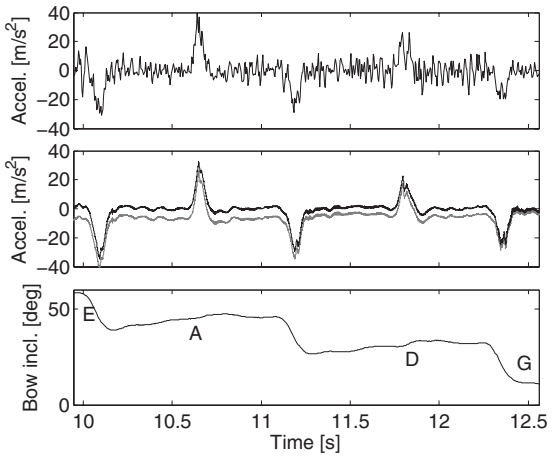


FIG. 6. Illustration of the correction of bow acceleration for inclination. Long sustained notes are played on the four strings (E A D G). The accelerometer is sensitive to bow inclination θ due to the influence of gravity. The inclination measured with the motion capture system was used for correcting the acceleration signal. Top: Acceleration obtained from motion capture data. Middle: Calibrated accelerometer signal without correction (gray) and with correction for inclination (black). Bottom: Bow inclination relative to the floor.

$$s = \frac{a_0(\theta)}{k} + s_0.$$

The calibrated bow acceleration a including the correction for inclination is then calculated as

$$a = k(s - s_0) - a_0(\theta).$$

The correction of the bow acceleration signal is illustrated in Fig. 6, showing sustained loud notes played on the four strings. The bow inclination (bottom) starts at a high value for the E string and approaches zero (horizontal) for the G string. The correction for inclination in the accelerometer data (middle) is about 7 $\mathrm{m/s^2}$ for the E string and almost zero for the G string.

For comparison, the bow acceleration as computed from the bow velocity measured with the motion capture system (low-pass filtered and cut-off of 18 Hz) is included (top). The loss in details is evident. In particular, the accelerometer signal provides more details at the peaks corresponding to the bow changes. The bow change actually gives rise to double acceleration peaks which merge in the smoothed acceleration signal provided by motion capture. The loss in detail is due to the combined effect of several factors, including a lower sampling frequency (250 Hz), the spatial resolution of the optical system, and the smoothing filter. The benefit of combining the motion capture system with an accelerometer is particularly obvious in cases like this, offering a higher time resolution during rapid changes.

B. Bow force

In Sec. IV B 1, a description of the calibration of the bow-force sensor and an assessment of the measurement accuracy will be provided. More details about the principles and design of the bow-force sensor can be found in Ref. 35.

1. Calibration procedure

The calibration of the bow-force sensor was performed using a calibrated load cell mounted on a small wooden board held by the player in a violin-like manner. The bow was held as normal with the hair flat on the load cell (no tilt) and perpendicular to the measuring direction of the cell. The player then pressed the bow against the load cell at ten equally spaced positions from the frog to the tip, applying a modulating force at each position.

The pose of the bow and the calibration board was tracked with the motion capture system, allowing for calculation of bow position and angles in a similar way as for the violin. The bow position was used for the calculation of the calibration coefficients. The measured bow angles can be used to check if the calibration was performed correctly as they might influence the quality of the calibration (see Sec. IV B 3). As will be shown below the calibration procedure needs to be repeated regularly during longer sessions as the calibration coefficients will change slightly over time due to changes in bow-hair tension (see Sec. IV B 2).

The calibration coefficients are obtained in two steps. In the first step, a second-order polynomial is fitted to the sensor response curves s_f (in volts) versus F_t (in newtons) for each of the ten fixed bow positions x_i [see examples in Fig. 7(a)],

$$s_f = b_2(x_i)F_t^2 + b_1(x_i)F_t, \tag{8}$$

yielding a set of coefficients b_2 (quadratic) and b_1 (linear). It should be noted that the constant term b_0 is not included in the equation as the sensor response s_f is compensated for the offset signal at zero bow force (referred to as "zero reference" further on).

Considering the mechanics of the bow a quadratic and linear dependence on bow position is expected for b_2 and b_1, respectively.[35] In the second step, curves are fitted to the obtained coefficients, yielding functional expressions of $b_2(x)$ and $b_1(x)$ dependent on bow position [see Figs. 7(b) and 7(c)].

Using the bow position $x=x_B$ obtained via motion capture the calibrated bow force can then be calculated from the sensor response s_f as

$$F_B(x) = \frac{-b_1(x) + \sqrt{b_1(x)^2 + 4b_2(x)s_f}}{2b_2(x)}. \tag{9}$$

The sensor response signal s_f is obtained by subtracting the offset voltage (zero reference) from the raw sensor signal. The zero reference can be determined at moments when the bow is off the string. The determination of the zero-reference points is done graphically in a semi-automatic procedure to obtain the highest possible accuracy (see Sec. IV B 2 for the influence of the zero reference on the accuracy of the force measurement). A zero-reference baseline is then constructed by interpolation.

A test of the calibration of bow force in a performance-like setting is shown in Fig. 8. A small wheel was mounted on the facing of the load cell, and short bow strokes were played on the wheel, using about 10 cm of the bow near the frog and near the tip, respectively. It can be seen that the

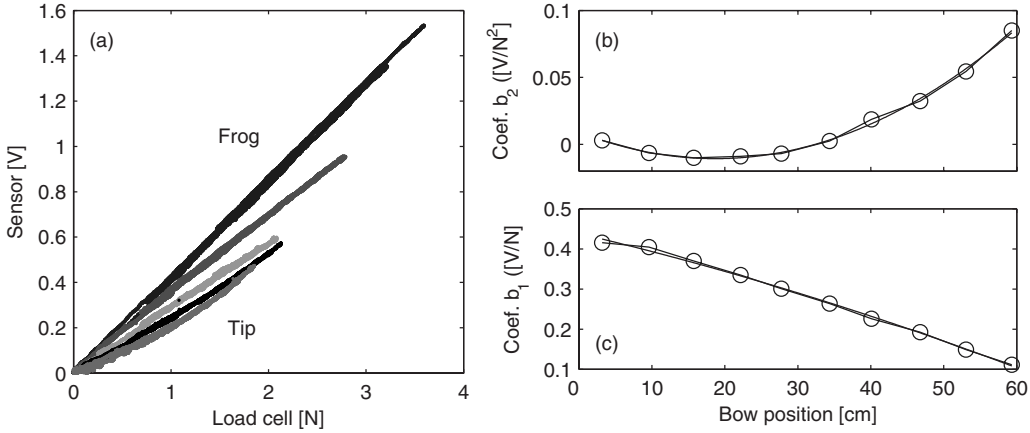


FIG. 7. The sensor is calibrated by pressing the bow hair on a load cell at ten equally spaced bow positions from the frog to the tip, applying a modulating force. (a) Sensor response s_f (offset subtracted) plotted versus reference force F_l measured by the load cell at fixed bow positions (five selected curves). (b) Quadratic and (c) linear coefficients b_2 and b_1 obtained from second-order polynomial fits for each bow position. The dependence of the coefficients on bow position can be described by second-order polynomials (fitted lines) used for calibration of bow force.

bow-force sensor showed a good correspondence with the reference force measured by the load cell, including a high degree of detail. The raw sensor signal (bottom panels) was clearly weaker at the tip, which was effectively compensated for by the calibration.

2. Accuracy of bow-force measurement

Two sources of error in the measured bow-force signal are as follows: (1) error in the calibration coefficients and (2) error in the determination of the zero reference of the raw sensor signal. The first can be due to the manual calibration procedure, including fluctuations in the orientation of the bow and the pressing direction on the load cell, or changes in bow-hair tension during playing under influence of humidity. Regarding the second error source, variations in the orientation of the bow can influence the camber of the bow stick due to gravitation and thus the offset deflection of the bow hair at the frog. Actually, such an influence was observed in the bow-force signal during the acceleration calibration measurements where the bow was slowly rotated in the air.

An assessment of the accuracy of the calibration was performed using the force calibrations from the experiment described in a companion paper.[32] Figure 9 shows an example of the evolution of the sensor response across seven calibration trials, roughly equally distributed over the session (about 2 h). It can be seen that there was a clear trend, which might be explained by a gradual change in bow-hair tension over the session. Irregularities in the manual calibration procedure would rather have resulted in random fluctuations and can therefore considered to have a minor influence in this case.

Another indication that the trend was the result of a change in bow-hair tension was that the zero reference also increased between subsequent calibrations. Both observations can be explained by a decrease in bow-hair tension resulting in the following: (1) the total stiffness of the bow decreases resulting in a larger deflection of the bow hair

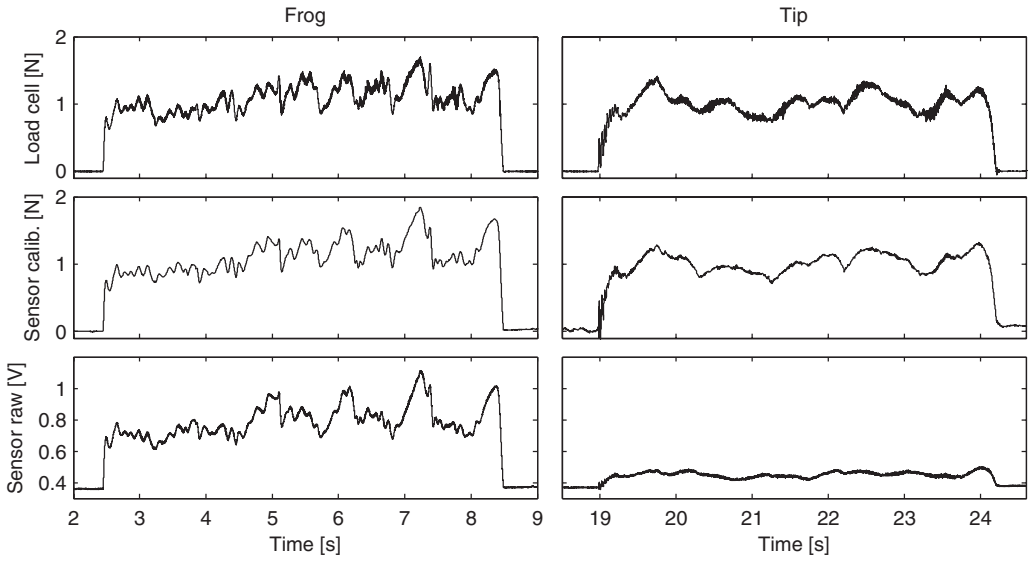


FIG. 8. Comparison of (top) reference signal measured by the load cell and (middle) bow-force signal measured by the sensor. The bow was moved back-and-forth on a small wheel mounted on the load cell near the frog (left) and near the tip (right). The bottom panels show the raw sensor signals.

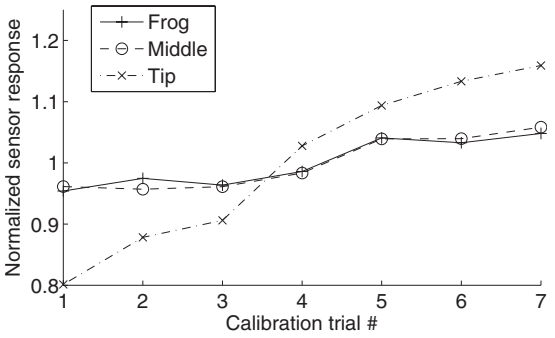


FIG. 9. Evolution of the sensor response over time during a measurement session (about 2 h) including seven calibration trials. The sensor response corresponding to a bow force of 1 N at the frog, middle, and tip was estimated using the calibration coefficients obtained by the force calibrations performed at regular intervals during the session. The sensor response was normalized by dividing with the average sensor response across trials.

under influence of loading and thus a larger sensor response, especially at the tip, and (2) the offset deflection of the bow-force sensor increases (the bow-force sensor is pressed against the bow hair causing an offset deflection at zero loading). In all measurements the correlation between variations in sensor response and zero reference was high ($r \approx 0.8-0.9$), indicating that bow-hair tension was the dominating factor in explaining the variations between calibrations.

From the above it can be concluded that the error in calibration should be considered locally in time. The following error analysis is therefore based on pairwise comparisons between subsequent calibrations. Because of the dependence of the sensor response on bow position and the non-linear behavior, especially at the tip, the error is considered at three bow positions (close to the frog, middle, and close to the tip) and five values of bow force representing the range of bow force in normal playing. For each pair of calibrations the normalized response difference $\Delta s_f/\bar{s}_f$ was calculated, and the average was taken across experimental sessions (three sessions, constituting a total of 20 calibrations with the same bow). The error was then transformed to force units (millinewtons) via the average response curves.

It was found that the calibration error was rather small with a relative error of about 2%, corresponding to an absolute error of about 20 mN at a bow force of 1 N. At low bow forces at the tip the error was slightly larger, up to 6%.

The second type of error, caused by uncertainty in the zero reference of the raw sensor signal, was assessed by considering the zero-reference fluctuations observed in the experimental data in Ref. 32. An (absolute) error in zero reference will have the largest relative influence at small bow forces, especially at the tip where the slope of the sensor response curve is shallow.

The combined estimated error due to calibration and uncertainty in zero-reference is shown in Table II. It can be observed that uncertainty in zero reference has a large impact at low bow forces, increasing to high levels at the tip. However, it should be realized that the estimated error presents a worst case scenario. The achieved accuracy is strongly dependent on the quality of the zero-reference signal, which can be improved by taking regular samples of the zero reference within recorded trials. This was taken into account in the design of the experimental tasks in the accompanying study[32] by inserting rests at regular intervals.

TABLE II. Estimated total error (worst case) due to calibration and zero-reference fluctuation in millinewtons and percent for the violin bow used in the study of Ref. 32.

F_t (N)	Frog	Middle	Tip
0.1	37 (37%)	49 (49%)	91 (91%)
0.5	38 (8%)	51 (10%)	68 (14%)
1.0	42 (4%)	54 (5%)	56 (6%)
1.5	47 (3%)	60 (4%)	51 (3%)
2.0	53 (3%)	66 (3%)	54 (3%)

3. Influence of bow tilt

In many situations players tilt the bow for obtaining a lighter contact with the hair on the string, especially close to the frog. As a result, the amount of bow hair in contact with the string is decreased, giving the player a finer control of bow force due to the increased effective compliance of the bow.

The force measured by the sensor is influenced by the bow tilt in two ways. First, the transversal force exerted by the bow on the string is no longer normal to the bow-force sensor. Second, there will be a force gradient across the width of the bow-hair ribbon, resulting in a non-uniform deformation of the leaf spring of the sensor. As a result, the measured bow force will deviate from the actual bow force.

For assessing the influence of bow tilt a test was performed, pressing the bow on the load cell with approximately constant force and slowly tilting the bow back-and-forth (about $\pm 30°$) at different bow positions. The discrepancy was largest close to the frog, where bow force measured by the sensor was underestimated by about 40% at a tilt angle of 30°. Toward the tip, the effect of the tilt diminished, giving deviations of about 20% at the middle and 5% at the tip.

The effect of tilting on measured bow force showed a regular behavior and can be compensated for effectively, using interpolation along a measured force calibration surface defined by bow position and tilt.[53]

V. DISCUSSION

A. Example

An illustration of the capability of the described system is given in Fig. 10 showing a complete set of bowing parameters in a violin solo performance. The combined panels provide a complete overview of the relevant aspects of bowing, including bow division (bow position—frog, middle, and tip), the use of the main bowing parameters (bow velocity, bow-bridge distance, bow force, and bow acceleration), and secondary control parameters (tilt and skewness of the bow). The bottom panel combines the extracted features "string played" and "bowing direction," showing the choices of the player and providing a rudimentary link to the score.

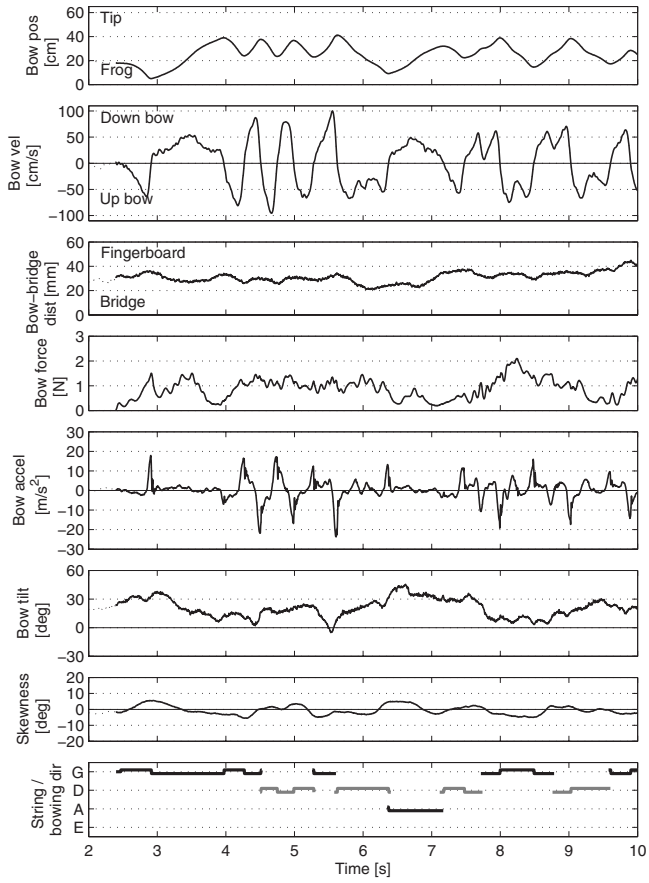


FIG. 10. Example of a full set of bowing parameters versus time measured with the described system. The musical fragment consisted of the first 1.5 bars of the Allemande of the second Partita for solo violin by J. S. Bach. The bowing parameters shown, from top to bottom, are *bow position*, *bow velocity*, *bow-bridge distance*, *bow force*, *bow acceleration*, the bow angles *tilt* and *skewness*, and the combined features *string played* (vertical position) and *bowing direction* (vertical offset, positive for up-bow and negative for down-bow).

The graphs contain detailed information about bowing patterns and coordination of bowing parameters in the performance. Both long notes (3–4 and 6.3–7.2 s) were played down-bow with crescendo-decrescendo. It can be clearly seen that the changes in bow velocity, bow force, and bow-bridge distance were coordinated to achieve the changes in dynamic level. Different bowing patterns, distinguishable in the example, are détaché (4–5.5 s) and portato (slurred pairs of notes, 7.5–9.7 s). The note at $t=8$ s (tonic) was stressed, which is clearly marked by the peak in bow force.

On a deeper level the use of bowing parameters can be analyzed in terms of the acoustics of bow-string interaction, for example, trajectories in the Schelleng diagram (bow force versus bow-bridge distance, steady state) or the Guettler diagram (bow force versus acceleration, transients). Analyses of an extensive collection of violin and viola performances using the described system will be reported in an accompanying study.[32]

B. Calculation of bowing parameters using kinematic models

In the described method the calculation of bowing parameters is based on fitted kinematic models of the violin and bow (position and orientation of segments), rather than on individual marker positions. This approach gives two important advantages: the calculated bowing parameters become less sensitive to noise and gaps in the marker trajectories, and the kinematic models allow for a general geometric definition of bowing parameters and bow angles, as described in Sec. III.

Even though the current work relied on kinematic modeling and fit facilities in the used Vicon software (Vicon iQ), the described methods can be generalized to other motion capture systems, both optical (active and passive) and electromagnetic tracking. For motion capture systems measuring 3D positions of markers, alternative kinematic fit methods could be used, e.g., the least-squares method described by Veldpaus *et al.*[44]

The kinematic models and calculations described in this study are specific to the violin and the viola. For measurement of cello and double bass bowing the same principles can be applied, but some adaptations are necessary due to the reversed orientation of the bow relative to the instrument.

C. Combination with sensors

In the current setup, the conditions for motion capture were not ideal. The camera distances were relatively large so that the full resolution of the cameras was not fully utilized for detailed measurement of bowing gestures. Furthermore, the recordings were made in a normal office environment, not optimized for motion capture. The reflections of the violin might have further degraded the performance of the motion capture system. The position data were therefore not suitable for the calculation of bow acceleration, and bow acceleration was measured with an accelerometer instead.

The complementary nature of motion capture techniques and the use of sensors attached to the bow offer several advantages, among others increased resolution in acceleration for transient analysis, practical and accurate calibration procedures for both acceleration and bow force, and correction of the influence of bow inclination on acceleration.

In an earlier study it has been shown that bow velocity profiles can be obtained from a combination of accelerometer signals and inexpensive motion tracking techniques based on normal video. The drift caused by integration can be limited by using the moments of detected bow changes as break points.[45] The same method can be applied using the measurements of the current system to obtain more detailed local estimates of bow velocity.

D. Measurement of bow force

The bow-force sensor in combination with the calibration procedure provided a reliable measurement of bow force, the estimated error being well below 10% at most combinations of bow position and bow force. The estimated error was largest at the tip at small bow forces, mainly due to the uncertainty in the zero reference (sensor offset voltage). It should be noted that the reported total error estimates represent a worst case scenario based on the average range of zero-reference fluctuations within recorded trials. The fluctuations in the zero reference can be effectively taken into

account by taking regular samples within recorded trials when the bow is off the string. The estimation of zero reference can be further improved by taking the orientation of the bow into account, compensating for the influence of gravity on the camber of the stick.

The bow force tended to be underestimated when the bow was tilted, especially close to the frog. This effect could also be compensated for via an extended calibration procedure, including bow tilt as a variable. However, the influence of tilt forms a fundamental problem in the measurement of bow force and will be further considered in future developments of the bow-force sensor.

E. Comparison with existing systems

As far as possible, comparisons with two recently developed systems for measurement of bowing gestures will be made: the hyperbow and the electromagnetic tracking method used by Maestre *et al.* (further referred to as the MTG method). Unfortunately, no extensive evaluations of these systems have been published, making direct comparisons difficult. The discussion below will therefore focus on fundamental aspects based on the descriptions of the measurement methods and inferences from reported results.

In the hyperbow[15] the orientation of the violin and the bow is measured using IMUs and the position of the bow relative to the violin by electric field sensing. The measurements were refined by Kalman filtering.

A fundamental problem with the use of IMUs for tracking position and orientation is that they measure second and first order derivatives (acceleration and angular velocity). Integration results in a cumulative error over time due to integration of noisy data (random walk) and gyrobias if no external measurement of position and orientation is available for adjustment.[46] The modified Kalman filter described in Ref. 15 operates on the derivatives of the quantities of interest (velocity and orientation), which implies that the drift problem was not effectively dealt with. The reported results in Ref. 15 show measurements of angular velocities rather than well-defined measurements of bow angles relative to the violin, indicating that the actual bow angles have not been calculated.

Regarding the measurement of bow position, a calibration graph comparing bow position obtained with motion capture and the Kalman-filtered electric field sensing from the hyperbow (Fig. 3.19 in Ref. 15) shows that the accuracy of the latter measurement would not meet the requirements set for our method. It can be concluded that optical motion capture measurements allow for a more straightforward and reliable extraction of bowing parameters and bow angles with a direct relevance to violin performance, which motivated our choice.

The electromagnetic tracking device used in the MTG method[29] forms an interesting alternative to optical motion capture, being more affordable and easier to use without the need of tedious post-processing for labeling markers and filling gaps. Another advantage is that such systems are better suited for real-time application. It should be noted that similar calibration methods as described in this study can be used for automatic feature extraction based on the geometry of the instrument. Possible accuracy issues are that the electromagnetic field is influenced by metal objects, and that the accuracy might degrade in rapid movements.[47,48] It remains to be evaluated how this affects the measurement of bowing parameters.

The measurement of bow force in the two systems discussed above is based on different principles. In the hyperbow, bow force is estimated from the change in camber at the middle of the bow stick using strain gauges. In the MTG method, measurement of bow force is based on motion capture without additional sensor. Bow force is measured as the distance between the straight lines corresponding to the string and the bow hair in case of no contact. When the bow is pressed into the string, this results in a discrepancy at the contact point due to the yielding of the bow hair and the string as well as the bending of the bow stick. The bow-force measurement was, however, not calibrated and therefore used as a relative indication of bow force only.[30,31] Both methods have in common that the signal is weakest close to the frog due to the limited bending of the stick, the sensitivity increasing toward the tip. This is in contrast to the bow-force measurement described in the current study, which is most sensitive close to the frog.

It should be noted that the force measurements in both the hyperbow and the MTG method are subject to the same fundamental effects of bow-hair tension and bow orientation as found in the assessment of bow-force measurement described in Sec. IV B 2, influencing the calibration and the zero offset. A standard calibration as used in Ref. 15 should therefore be used with caution, carefully controlling the tension of the bow hair. Furthermore, the variations in the offset of the sensor signal should be included in the calculation of bow force.

The force calibration measurements in this study were used for a preliminary comparison between our sensor-based measurement of bow force and a reconstruction of bow force from motion capture data inspired by the MTG method. It was found that a calibration of the latter could be provided for measurement of bow force in newtons. The motion capture method gave reasonable force reconstructions in the upper half of the bow where the bending of the stick under influence of bow force is considerable. In the lower half the reconstruction became less reliable, mainly because the change in bow-string distance due to bending of the bow is much smaller and harder to estimate reliably. Further complications are due to the deflection of the string under influence of bow force and stopping of the string. These deviations are, however, small and close to the spatial resolution of the motion capture system and therefore hard to take into account. Compared to the sensor-based measurement of bow force the motion capture method resulted in a rather noisy signal and provided less detail.

A drawback of the currently described setup is that it is confined to the laboratory. However, for the purposes of the accompanying study[32] this was an acceptable limitation.

F. Applications

The methods for determination of bowing parameters developed in this study allow for detailed measurements of bowing gestures and in-depth analyses of players' control and coordination strategies in performance. In addition, motion capture can be extended to include body movements of the player, allowing study of biomechanical aspects of playing, anticipatory movements, and the influence of posture. This type of studies will deepen our understanding of bowed-string instrument performance, not only from a scientific point of view but also from the player's perspective. Especially in combination with effective visualizations, motion capture and other measurement techniques have interesting potential for pedagogical use.[2,27,28,49,50] Another promising application is gesture-based sound synthesis, where this type of measurements will be helpful in the development of parametric control models.[30,51–53]

VI. CONCLUSIONS

In this study motion capture techniques have been applied to build a complete system for measurement of bowing parameters in violin playing. Kinematic models for the violin and bow have been developed which can be used to calculate the main bowing parameters (bow position, bow velocity, and bow-bridge distance). In addition, the angles of the bow relative to the violin (inclination, skewness, and tilt) which have a limited direct influence on the generated sound, but reflect the pre-planning and coordination of parameters in bowing gestures, are measured.[54] Bow force is measured with a custom-designed sensor, integrated with the frog, allowing for accurate measurement of bow force in combination with motion capture data. Bow acceleration is measured in two directions with an accelerometer mounted on the frog, allowing for detailed study of transients during bow changes or attacks.

In addition, the system allows automatic extraction and visualization of three basic performance features: bowing direction, moments of on/off in bow-string contact, and determination of which string is played. These features are necessary for musically relevant analyses of bowing parameter data. Such analyses will give an understanding of how a certain bowing gesture is composed of basic elements of motion and allow comparison of bow control strategies between players.

The described system is capable of measuring all bowing parameters in violin performance without disturbing the player. The system is robust, and accuracy and signal quality are high. In combination with the extracted features, the system allows for detailed studies of musically relevant aspects of bow control and coordination of bowing parameters in bowed-string instrument performance.

ACKNOWLEDGMENTS

This work was partly supported by the Swedish Science Foundation (Contract No. 621-2001-2537) and the Natural Sciences and Engineering Research Council of Canada (NSERC-SRO). The bow-force sensor was developed as part of the CONSONNES project funded by the French Agence Nationale pour le Recherche. The collaboration was stimulated by two short-term scientific missions (E.S. at IRCAM, Paris and M.D. at IDMIL, McGill University, Montreal) funded by the Cost 287-ConGas action. Important part of this work was realized at IDMIL, McGill University, Montreal. Part of this work was performed during the second author's (M.D.) stay at the Department of Speech, Music and Hearing, Royal Institute of Technology (KTH), Stockholm supported by the Swedish Institute. Special thanks go to Emmanuel Fléty and Alain Terrier for their valuable help in developing the force sensor, to Nicolas Rasamimanana and Frédéric Bevilacqua with whom this work was initiated, to Marcelo M. Wanderley for his inspiring support and advice, and to Lambert Chen for his invaluable contribution as a violin and viola player.

[1]W. Goebl and R. Bresin, "Measurement and reproduction accuracy of computer-controlled grand pianos," J. Acoust. Soc. Am. **114**, 2273–2283 (2003).
[2]P. Hodgson, *Motion Study and Violin Bowing* (American String Teachers Association, Urbana, IL, 1958), first published in 1934 by J. H. Lavender & Co., London.
[3]A polemic review of Hodgson's book was published in The Musical Times (Ref. 55), followed by a reaction of the author (Ref. 56).
[4]A. Askenfelt, "Measurement of bow motion and bow force in violin playing," J. Acoust. Soc. Am. **80**, 1007–1015 (1986).
[5]A. Askenfelt, "Measurement of the bowing parameters in violin playing. II: Bow-bridge distance, dynamic range, and limits of bow force," J. Acoust. Soc. Am. **86**, 503–516 (1989).
[6]T. Machover, "Hyperinstruments—A progress report 1987–1991," MIT Technical Report, MIT Media Laboratory, Cambridge, MA (1992).
[7]D. L. Trueman, "Reinventing the violin," Ph.D. thesis, Princeton University, Princeton, NJ (1999).
[8]C. Nichols, "The vBow: A virtual violin bow controller for mapping gesture to synthesis with haptic feedback," Organised Sound **7**, 215–220 (2002).
[9]D. Young, "New frontiers of expression through real-time dynamics measurement of violin bows," MS thesis, Massachusetts Institute of Technology, Cambridge, MA (2001).
[10]D. Young, "The Hyperbow controller: Real-time dynamics measurement of violin performance," in Proceedings of the NIME-02 Conference on New Instruments for Musical Expression (2002).
[11]D. Young, P. Nunn, and A. Vassiliev, "Composing for Hyperbow: A collaboration between MIT and the Royal Academy of Music," in Proceedings of the 2006 International Conference on New Interfaces for Musical Expression (NIME06) (2006).
[12]D. Overholt, "The overtone violin," in Proceedings of the 2005 International Conference on New Interfaces for Musical Expression (NIME05), Vancouver, BC, Canada (2005), pp. 34–37.
[13]F. Bevilacqua, N. Rasamimanana, E. Fléty, S. Lemouton, and F. Baschet, "The augmented violin project: Research, composition and performance report," in Proceedings of the 6th International Conference on New Interfaces for Musical Expression (NIME06), Paris, France (2006).
[14]M. T. Marshall, J. Malloch, and M. M. Wanderley, "Non-conscious control of sound spatialization," in Proceedings of the 4th International Conference on Enactive Interfaces (ENACTIVE07), Grenoble, France (2007), pp. 377–380.
[15]D. Young, "A methodology for investigation of bowed string performance through measurement of violin bowing technique," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA (2007).
[16]J. A. Paradiso and N. Gershenfeld, "Musical applications of electric field sensing," Comput. Music J. **21**, 69–89 (1997).
[17]S. Serafin and D. Young, "Bowed string physical model validation through use of a bow controller and examination of bow strokes," in Proceedings of the Stockholm Music Acoustics Conference (SMAC 03), Stockholm, Sweden (2003).
[18]N. Rasamimanana, "Gesture analysis of bow strokes using an augmented violin," MS thesis, Université Pierre et Marie Curie, Paris VI, France (2003).
[19]N. Rasamimanana, E. Fléty, and F. Bevilacqua, "Gesture analysis of violin

bow strokes," in *Gesture in Human-Computer Interaction and Simulation: 6th International Gesture Workshop, Revised Selected Papers* (Springer, Berlin, 2006), pp. 145–155.
[20]N. Rasamimanana, "Geste instrumental du violoniste en situation de jeu: Analyse et modélisation (Violin player instrumental gesture: Analysis and modelling)," Ph.D. thesis, Université Pierre et Marie Curie (UPMC), Paris VI, France (2008).
[21]A. P. Baader, O. Kazennikov, and M. Wiesendanger, "Coordination of bowing and fingering in violin playing," Brain Res. Cognit. Brain Res. **23**, 436–443 (2005).
[22]H. Winold, E. Thelen, and B. D. Ulrich, "Coordination and control in the bow arm movements of highly skilled cellists," Ecological Psychol. **6**, 1–31 (1994).
[23]G. Shan and P. Visentin, "A quantitative three-dimensional analysis of arm kinematics in violin performance," Med. Probl. Perform. Art. **18**, 3–10 (2003).
[24]P. Visentin and G. Shan, "The kinetic characteristics of the bow arm during violin performance: An examination of internal loads as a function of tempo," Med. Probl. Perform. Art. **18**, 91–97 (2003).
[25]L. Turner-Stokes and K. Reid, "Three-dimensional motion analysis of upper limb movement in the bowing arm of string-playing musicians," Clin. Biomech. (Bristol, Avon) **14**, 426–433 (1999).
[26]C. Peiper, D. Warden, and G. Garnett, "An interface for real-time classification of articulations produced by violin bowing," in Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME-03), Montreal, QC, Canada (2003).
[27]F. Rabbath and H. Sturm, "Art of the bow with Francois Rabbath," http://www.artofthebow.com (Last viewed 7/1/2009).
[28]K. Ng, O. Larkin, T. Koerselmans, B. Ong, D. Schwartz, and F. Bevilacqua, "The 3D augmented mirror: Motion analysis for string practice training," in *Proceedings of the 2007 International Computer Music Conference (ICMC07)* (The International Computer Music Association, San Francisco, CA, 2007), Vol. **II**, pp. 53–56.
[29]E. Maestre, J. Bonada, M. Blaauw, A. Pérez, and E. Guaus, "Acquisition of violin instrumental gestures using a commercial EMF tracking device," in *Proceedings of the 2007 International Computer Music Conference (ICMC07)* (The International Computer Music Association, San Francisco, CA, 2007), Vol. **I**, pp. 386–393.
[30]A. Perez, J. Bonada, E. Maestre, E. Guaus, and M. Blaauw, "Combining performance action with spectral models for violin sound transformation," in Proceedings of the 19th International Congress on Acoustics (ICA07), Madrid, Spain (2007).
[31]A. Perez, J. Bonada, E. Maestre, E. Guaus, and M. Blaauw, "Score level timbre transformations of violin sounds," in Proceedings of the 11th International Conference on Digital Audio Effects (DAFx-08), Espoo, Finland (2008).
[32]E. Schoonderwaldt, "The player and the bowed string: Coordination of bowing parameters in violin and viola performance," J. Acoust. Soc. Am. **126**, 2709–2720 (2009).
[33]http://www.vicon.com (Last viewed 7/1/2009).
[34]Alternatively, digital probing techniques could be used, e.g., as described in Ref. 29.
[35]M. Demoucron, A. Askenfelt and R. Causse, "Measuring bow force in bowed string performance: Theory and implementation of a bow force sensor," Acta Acust. Acust. **95**, 718–732 (2009).
[36]A. Askenfelt, "Observations on the violin bow and the interaction with the string," in Proceedings of the International Symposium on Musical Acoustics (ISMA95), Dourdan, France (1995), pp. 197–212.
[37]N. Pickering, "Physical properties of violin bows," J. Violin Society of America **8**, 41–57 (1987).
[38]http://www.mathworks.com (Last viewed 7/1/2009).
[39]For an extensive overview of conversions between different representations of orientation, see Ref. 40.
[40]J. Diebel, "Representing attitude: Euler angles, unit quaternions, and rotation vectors," Stanford University Technical Report, Stanford University, Stanford, CA, 2006.
[41]To obtain the acoustically more relevant *relative* bow-bridge distance β, y_{bb} should be divided by the effective length of the string. This requires knowledge of the position where the string is stopped by the player, which could be obtained, e.g., by pitch analysis of the audio recording.
[42]To avoid possible confusion around this subject it should be noted that the bow angles as defined here do not exactly equal the Tait–Bryan (or Euler) angles of the relative orientation of the bow and the violin, as suggested in Ref. 43.
[43]E. Schoonderwaldt, S. Sinclair, and M. Wanderley, "Why do we need 5-DOF force feedback? An analysis of violin bowing," in Proceedings of the ENACTIVE07 (2007).
[44]F. E. Veldpaus, H. J. Woltring, and L. J. M. G. Dortmans, "A least-squares algorithm for the equiform transformation from spatial marker coordinates," J. Biomech. **21**, 45–54 (1988).
[45]E. Schoonderwaldt, N. Rasamimanana, and F. Bevilacqua, "Combining accelerometer and video camera: Reconstruction of bow velocity profiles," in Proceedings of the 6th International Conference on New Interfaces for Musical Expression (NIME06), Paris, France (2006), pp. 200–203.
[46]A. Y. Benbasat, "An inertial measurement unit for user interfaces," MS thesis, Massachusetts Institute of Technology, Cambridge, MA (2000).
[47]D. D. Frantz, A. D. Wiles, S. E. Leis, and S. R. Kirsch, "Accuracy assessment protocols for electromagnetic tracking systems," Phys. Med. Biol. **48**, 2241–2251 (2003).
[48]N. B. Schuler, M. J. Bey, J. T. Shearn, and D. L. Butler, "Evaluation of an electromagnetic position tracking device for measuring in vivo, dynamic joint kinematics," J. Biomech. **38**, 2113–2117 (2005).
[49]T. K. Ho, "A computer-assisted approach to the teaching of violin tone production," ACM SIGCUE Outlook **21**, 73–83 (1991).
[50]E. Schoonderwaldt and M. M. Wanderley, "Visualization of bowing gestures for feedback: The Hodgson plot," in Proceedings of the 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution (AXMEDIS07) (2007), Vol. **II**, pp. 65–70, for examples, see http://www.youtube.com/schoondw (Last viewed 7/1/2009).
[51]M. Demoucron and R. Caussé, "Sound synthesis of bowed string instruments using a gesture based control of a physical model," in Proceedings of the International Symposium on Musical Acoustics (ISMA07) (2007).
[52]D. Young and S. Serafin, "Investigation the performance of a violin physical model: Recent real player studies," in *Proceedings of the 2007 International Computer Music Conference (ICMC07)* (The International Computer Music Association, 2007), Vol. **I**, pp. 394–397.
[53]M. Demoucron, "On the control of virtual violins: Physical modelling and control of bowed string instruments," Ph.D. thesis, Université Pierre et Marie Curie (UPMC), Paris, France and Royal Institute of Technology (KTH), Stockholm, Sweden (2008).
[54]Arguably, tilt has a certain influence on the spectrum, as shown in Ref. 57, but the effect is small compared to that of, e.g., bow force.
[55]F.B., "Motion study and violin bowing. By Percival Hodgson," The Musical Times **76**, 131 (1935), anonymous review of Hodgson's book "Motion study and violin bowing" (Ref. 2).
[56]P. Hodgson, "Motion study and violin bowing," The Musical Times **76**, 347–348 (1935).
[57]E. Schoonderwaldt, K. Guettler, and A. Askenfelt, "Effect of the width of the bow hair on the violin string spectrum," in Proceedings of the Stockholm Music Acoustics Conference (SMAC03), Stockholm, Sweden (2003), pp. 91–94.

The player and the bowed string: Coordination of bowing parameters in violin and viola performance

E. Schoonderwaldt[a)]
Department of Speech Music and Hearing, School of Computer Science and Communication, KTH, Lindstedtsvägen 24, SE-100 44 Stockholm, Sweden and Input Devices and Music Interaction Laboratory, Schulich School of Music, McGill University, 555 Sherbrooke Street West, Montreal, Quebec H3A 1E3, Canada

(Received 17 March 2009; revised 6 July 2009; accepted 7 July 2009)

An experiment was conducted with four violin and viola players, measuring their bowing performance using an optical motion capture system and sensors on the bow. The measurements allowed for a detailed analysis of the use and coordination of the main bowing parameters bow velocity, bow force, and bow-bridge distance. An analysis of bowing strategies in détaché playing of notes of three durations (0.2, 2, and 4 s) at three dynamic levels (*pp*, *mf*, and *f*) on all four strings is presented, focusing on the "steady" part of the notes. The results revealed clear trends in the coordinated variations of the bowing parameters depending on the constraints of the task, reflecting a common behavior as well as individual strategies. Furthermore, there were clear indications that the players adapted the bowing parameters to the physical properties of the string and the instrument, respecting the limits of the playable control parameter space.
 [DOI: 10.1121/1.3203209]

PACS number(s): 43.75.De [NHF] Pages: 2709–2720

I. INTRODUCTION

A. General background

The violin is by many considered as one of the most expressive musical instruments, offering a rich variety of tone color and allowing for an almost unlimited degree of expression and virtuosity in the hands of an achieved performer. At the same time, the violin is known to be one of the most difficult instruments to master, and it takes some skill even to produce a simple tone.

Tone production in violin playing involves many constraints. From an acoustical perspective the production of a good tone requires a subtle coordination between the main bowing parameters bow velocity, bow force, and bow-bridge distance. From a motor control perspective, bowing requires a precise control and coordination of the different parts of the bowing arm, which is only achieved after extensive training.[1,2] All these aspects are subordinate to the music to be performed, which can therefore be considered as a third type of constraint.

The conditions for the production of an acceptable tone have already a long history of research, beginning with the mechanical bowing experiments by Raman,[3] leading to a theoretical description of the constraints for steady-state Helmholtz motion by Schelleng.[4] These constraints are known as the Schelleng equations for maximum and minimum bow force. In the so-called Schelleng diagram, a log-log representation of bow force versus bow-bridge distance at constant bow velocity, the bow-force limits form straight lines, demarcating a triangular-shaped playable region.

Similarly, the conditions for the creation of Helmholtz motion during attacks were formalized by Guettler[5] in terms of bow force, bow acceleration, and bow-bridge distance. The conditions for a perfect attack, characterized by Helmholtz motion right from the start, are confined to a triangular-shaped region in the Guettler diagram of bow force versus bow acceleration at constant bow-bridge distance. Both the Schelleng and the Guettler diagram provide a macroscopic description of the dependence of string vibrations on a limited set of control parameters, and have therefore become a hallmark of playability evaluations of physical bowed-string models, allowing for comparison with realistic measurements and study of the effect of different friction models and other synthesis parameters.[6,7]

Measurements of bowing parameters in real violin performance were undertaken by Askenfelt[8,9] in two pioneering studies. These studies provided a first insight in the use and coordination of the main bowing parameters under a variety of conditions, including dynamic level, note duration, and typical bowing techniques such as spiccato and martellato. The measured bowing parameters proved also useful for analysis of timing and phrasing in musical fragments as well as the expression of mood, clearly revealing the means used by the players to transform their intentions into sounding results. The bowing parameters showed a clear relation with the vibration level of the violin, measured by a contact microphone on the top plate. Moreover, it was shown that the violinists kept a safe margin with respect to the Schelleng bow-force limits in the production of sustained notes, mov-

[a)]Author to whom correspondence should be addressed. Electronic mail: schoondw@kth.se

ing more or less in parallel with the bow-force limits through the parameter space when playing at different dynamic levels.

More recently, the interest in the measurement of bowing gestures has been revived in connection with augmented interactive performance and the control of virtual violins.[7,10–13] Physical models are bound to similar constraints of the coordination of bowing parameters as real violins. For example, the quality of the attack is strongly dependent on detailed features of the acceleration and force envelopes.[13] Moreover, a realistic synthesis of violin sound, including the reproduction of a variety of bowing techniques, can only be achieved when the control parameters are varied in a way resembling a real violin performance. Real-time controllers, as well as score-based synthesis methods, must be able to produce realistic control envelopes, and their design therefore requires an in-depth understanding of the mechanics and acoustics of bowing, along with a detailed insight in typical bowing strategies in a variety of musical contexts.

B. Aims of the study

The major goal of the current study is to provide a detailed description of the interaction between the player and the instrument and to show how the player adapts to the constraints imposed by the bow-string interaction. In earlier studies it has been shown that players generally respect the Schelleng limits of bow force.[9,10] However, no detailed studies are available of how players adapt the bowing parameters to the physical properties of, e.g., different strings. The degree in which the player adapts to the varying physical constraints will therefore provide interesting new insights in the sensitivity of the player with respect to the playability of bowed-string instruments.

A method developed for measuring bowing parameters developed by Schoonderwaldt and Demoucron[14] was used for an extensive study of violin and viola performances. The experiment covered a wide variety of bowing techniques, both isolated in basic performance tasks and applied in a proper musical context. The measurements represent an inventory of typical aspects of bowing technique and should provide a detailed insight in the coordination and control of bowing parameters by players under realistic performance conditions.

The analyses in this paper will be restricted to the "steady" part of bowing, focusing on aspects of coordination and control in sustained notes. The study consists of three parts. In the first part the coordination of bowing parameters will be analyzed in relation with the bow-force limits in the Schelleng diagram. The chosen combinations of bowing parameters by the participating players can be directly compared to the bow-force limits found in an earlier empirical study using a bowing machine.[15] Bowing strategies for producing different dynamic levels will be analyzed dependent on note duration, shedding further light on the trade-off between bow velocity and bow-bridge distance in setting (or changing) the dynamic level, as suggested by Askenfelt.[9] Further, individual patterns in the ranges and variation of bowing parameters will be identified and compared to the common strategies.

In the second part, the dependence of sound level and spectral centroid on the main bowing parameters will be analyzed and compared to findings of an earlier study using a bowing machine.[16] The results provide further insight in the control of these sound features in terms of the main bowing parameters, and give an indication of the typical ranges found in violin performance.

In the third part subtle aspects of control exerted by players will be studied by analyzing how the participants adapted their combinations of bowing parameters to the physical properties of the string and the instrument. Within the same instrument the physical properties of the strings, in particular the characteristic impedance and internal damping, differ significantly from the lowest to the highest strings, giving rise to, among others, different bow-force limits. Such differences are even more pronounced between different members of the violin family, for example, the violin and the viola. In the latter case the difference in size of the instruments also requires an adaptation of posture, which presents an additional difficulty for a player when switching from one instrument to another.

II. METHOD

A. Setup

In the experiment, motion data, sensor signals, and sound were synchronously recorded. The position and orientation of the bow and the violin/viola were tracked using a Vicon 460 optical motion capture system. The bow was equipped with a bow-force sensor and an accelerometer, mounted on the frog. A detailed specification of the setup, the calibration steps involved, and the definitions and calculations of the bowing parameters is provided in a companion paper.[14]

B. Participants

A total of six violin and/or viola players participated in the study. All players were advanced master and postgraduate students from two schools of music in Montreal [Schulich School of Music at McGill University and the faculty of music at the Université de Montréal (UdeM)]. Two of the players performed on both the violin and the viola, making a total of eight recording sessions. All players were paid 20 Canadian dollars in compensation.

For the current analyses two sessions were discarded, due to problems with the calibration of the bow force in the upper half of the bow. A total of six sessions remained, three with violin and three with viola. An overview of the sessions and participants included in the analyses is shown in Table I.

C. Experimental procedure and tasks

During the experiment the players were seated on a piano stool. The wires from the sensors on the bow were taped to the right lower and upper arm to avoid interference in playing, taking care that there was ample freedom for the

TABLE I. Overview of participants included in the analyses. Players P4 and P5 performed on violin and viola, yielding a total of six recording sessions.

Player ID	Sex	Violin	Viola	Details
P3	f		√	Master student viola (final year)
P4	m	√	√	Master student viola (final year)
P5	m	√	√	Post-graduate student viola
P6	f	√		Master student (final year)

performance of full bow strokes. After these preparations the players were given time to familiarize with the experimental situation, making sure that they could play comfortably.

The same instrument and bow combinations were used in all recording sessions, with exception of player P3, who played on a smaller viola. The provided bows and instruments were of master quality.

Before the start of the experiment, the participants received general oral instructions regarding the details of the experimental procedure and the calibration of the bow-force sensor. Force calibration was performed at regular intervals during a session (six to seven times) by the players themselves.[14]

A variety of bowing techniques were recorded, including détaché (slow and fast), spiccato, martelé, tremolo, and different types of attacks. The tasks consisted of basic tasks (repeated notes on different strings and at different dynamic levels), scales, and musical excerpts, all presented in normal musical notation. The different types of tasks were intertwined so that the basic tasks were followed by the application of the same bowing technique in a musical context, which helped to keep the participants motivated. All participants performed the tasks in the same order. The sessions lasted typically 2 h, including two 5–10 min breaks.

Explicit reference to bowing parameters was avoided in the instructions, leaving the decisions up to the players. The players did not receive feedback about their performance; only in exceptional cases feedback was given, for example, when the dynamic contrast was judged inappropriate.

The basic tasks were performed in strict tempo indicated by a metronome signal, presented to the players' right ear via light earplug headphones. The metronome signal was recorded on a separate audio track. For the musical excerpts the target tempo was indicated by a two-bar cue from the metronome via a loudspeaker.

The basic tasks were performed on all strings, stopping the string with the third finger in first position, a musical fourth above the open string (C4, G4, D5, and A5 on the violin G, D, A, and E strings, stopped string length 244 mm). The players were allowed to change fingering to prevent fatigue. For viola the tasks were similar, transposed a fifth down (stopped string length 282–285 mm).

The tasks selected for analysis consisted of sustained notes (whole notes and half notes) and détaché 16th notes, played at 3 dynamic levels (forte, mezzoforte, and pianissimo), as well as 4 half-note conditions with various crescendo-diminuendo patterns (between notes and within notes). The nominal durations of the whole and half notes were 4 and 2 s (metronome at 60 BPM). For the 16th notes the nominal duration was 0.2 s (76 BPM). The long-note conditions consisted of 4 notes per string and dynamic level, and the 16th-note conditions of 24 notes.

D. Analysis

In the following analyses, only the steady parts of the notes were considered. Samples were collected from the separate trials by an automatic procedure, respecting a certain margin before and after bow changes (200 and 50 ms in long notes and 16th notes, respectively). Another requirement was that the bow force should exceed a minimum threshold of 0.01 N, to make sure that the bow was in contact with the string. The collected points were more sparsely sampled than the original data, with effective sample rates of 12.5 and 50 Hz for long notes and 16th notes, respectively. The effective playing time accounted for in the analyses per player, string, and dynamic level amounted to 15, 7, and 2.5 s for the whole-note, half-note, and 16th-note conditions, respectively, corresponding to about 188, 88, and 120 samples.

The accumulated data of all participants per instrument (violin/viola) included 144 notes for each of the long-note conditions (48 per dynamic level), and 864 notes (288 per dynamic level) for the 16th-note conditions. The effective playing times considering only the steady part of the notes were about 540, 250, and 86 s for the whole-note, half-note, and 16th-note conditions, respectively.

The bowing parameters included were bow position, bow velocity, bow-bridge distance (absolute and normalized with respect to effective string length), and bow force. The sound features included sound level (rms converted to decibel) and spectral centroid (compensated for background noise). All data points were labeled with respect to instrument (violin/viola), player, string played, note duration, dynamic level, and bowing direction.

The recording level of the sound was not calibrated. The sound level of the violin performances was therefore post-hoc compensated for the level difference between recording sessions. The differences were estimated by comparing the sound levels in specific central regions of the bowing-parameter spaces in the sustained part of long notes. This method of compensation was judged appropriate for the purpose of the current study by the apparent predictability of sound level by the v_B/β ratio (see Sec. IV A).

III. COORDINATION OF BOWING PARAMETERS IN THE SCHELLENG DIAGRAM

Typical examples of bowing-parameter signals as a function of time are shown in Fig. 1. The example shows performances of the basic tasks with three different note lengths (whole notes, half notes, and 16th notes) played *mf*. The bowing parameters shown from the top down are bow position (x_B), bow velocity (v_B), bow force (F_B), and relative bow-bridge distance (β). The following analyses will mainly involve distributions of the latter three main bowing parameters.

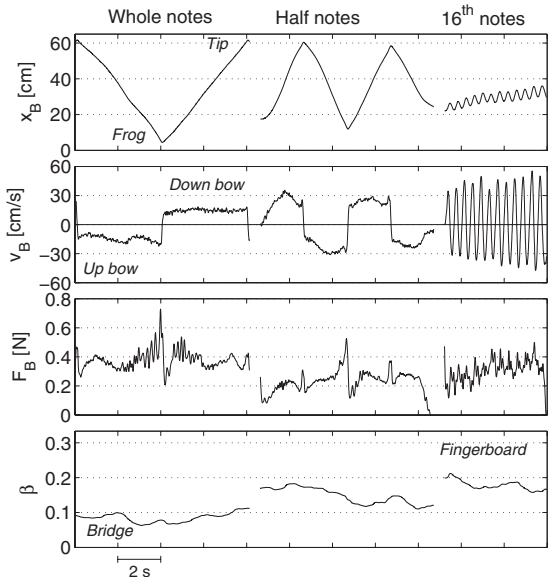


FIG. 1. Examples of the analyzed bowing-parameter signals of the three note-length conditions: whole notes, half notes, and 16th notes. (Data: violin, player P4, D string, and *mf*.)

A. Sustained notes

Figure 2 shows the density distributions of bow force and relative bow-bridge distance when the players performed on the violin D string, represented in Schelleng diagrams at six adjacent ranges of bow velocity. The shown data provide a global representation of all dynamic levels in the whole-note and half-note conditions by the three violin players (P4, P5, and P6). The densities of the respective note-length conditions were normalized, ensuring an equal representation in the graphs. The limits of the playable region were estimated for the same violin and same type of string (Pirastro Obligato, D string stopped at pitch G4), with the use of a bowing machine. Additional information regarding the upper and lower bow-force limits is based on the findings by Schoonderwaldt *et al.*[15] (see figure caption and footnote[17] for more details).

The combined data from the three players formed clear coherent regions in the Schelleng diagrams, mostly following the contours of the indicated bow-force limits. The highest densities of points were found at bow velocities around 15 and 30 cm/s, corresponding to the whole- and half-note conditions, respectively. As the individual players used somewhat different ranges in bowing parameters (see Sec. III D), the distinction between dynamic levels has become somewhat blurred. However, in panel (c) ($v_B \approx 15$ cm/s), the three dynamic levels can still be distinguished as more or less separate clusters.

In general, the upper bow-force limit was respected with a reasonable margin, especially at higher bow velocities, and no occurrences of raucous motion could be heard in the sound recordings. Furthermore, it could be observed that higher bow forces were used at higher bow velocities, in accordance with the increase in the upper bow-force limit. In some cases at forte levels the pitch-flattening limit indicated

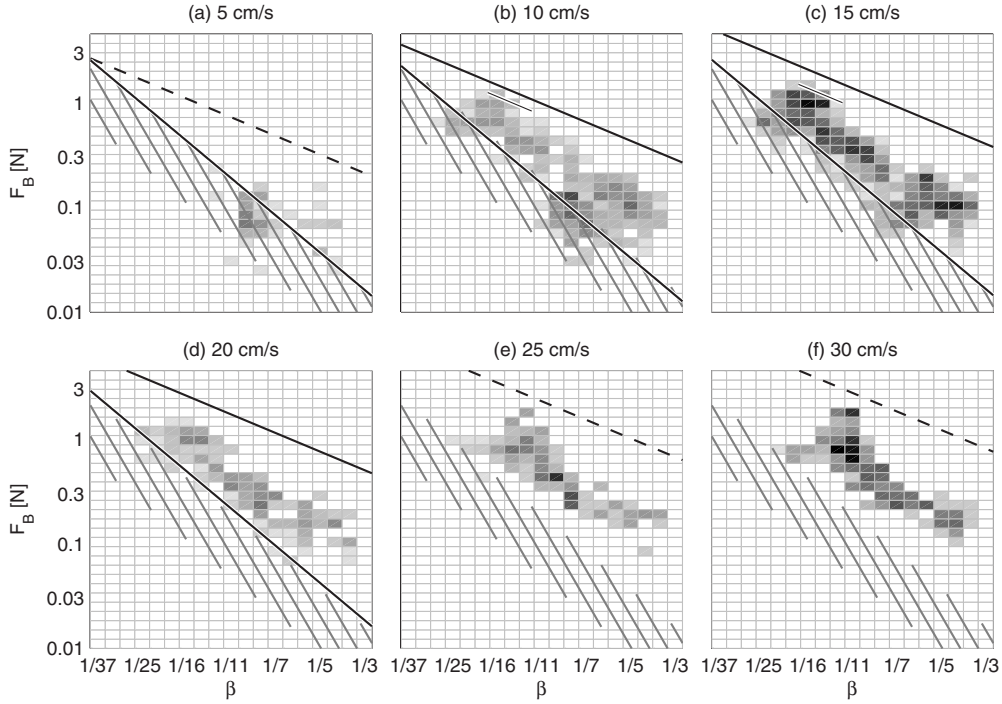


FIG. 2. Density distributions of bow force and bow-bridge distance represented in Schelleng diagrams at six adjacent ranges of bow velocity, for whole and half notes performed by the three violin players on the violin D string. The bow-velocity range per panel was 5 cm/s wide, centered around the indicated bow velocities. The density is indicated by the gray scale (the higher the density, the darker). The indicated limits of the playable region are based on measurements with a bowing machine completed with findings from an earlier study (Ref. 15). The solid lines represent the fitted bow-force limits measured for a stopped Pirastro Obligato D string (pitch: G4) mounted on the violin used in the performances. The bow force limits were not measured at all bow velocities; in those cases the upper bow-force limit is indicated by dashed lines, based on extrapolation or representative measurements from an earlier study (Ref. 15). Additionally, in panels (b) and (c), the 10 cent pitch-flattening limits, measured for the used instrument-string combination for a stopped note (G4), are indicated by short solid lines at about $\beta = 1/16$. The hatched areas indicate the range of lower bow-force limits found earlier (Ref. 15).

by the short solid lines in panels (b) and (c) was exceeded, which was indeed audible in the recorded sound as a "pressed" tone with instable pitch.

As pointed out by Schoonderwaldt *et al.*,[15] there was a higher degree of uncertainty associated with the determination of the lower bow-force limit, which was shown to be strongly influenced by finger damping. As this was an uncertain factor in the current experiment, the location of the lower bow-force limit could not be exactly estimated. The lower bow-force limit is therefore indicated by a band, representing the range of lower limits found under different conditions reported in that study.[15] Another assumption based on the earlier findings is that the lower limit is independent of bow velocity, in contrast to Schelleng's predictions. This behavior is extrapolated to bow velocities of 25 and 30 cm/s, not measured in the earlier study.

At the three highest bow velocities in Fig. 2 the data points fell mostly above the indicated lower bow-force limits. In contrast, at lower bow velocities, most notably 5 and 10 cm/s, combinations of bow force and β below the measured lower bow-force limit were observed, and there was a considerable overlap with the observed playing regions and the hatched area representing the possible transition to multiple slipping. In particular, for the three lowest bow velocities the data points for bow forces lower than 0.1 N and β values between 1/11 and 1/7 fell almost entirely in the indicated gray zone, and could be mainly attributed to the *pp* whole-note condition by one of the players. However, no prolonged episodes of multiple slipping were audible in the recorded sound, with some exceptions in the vicinity of some bow changes. More generally, it could be concluded that the displayed points in Fig. 2 were mostly associated with regular Helmholtz motion.

B. Extension to higher bow velocities

In normal violin performance, bow velocities of 1 m/s are not exceptional and bow velocities beyond 2 m/s can occasionally be observed. Usually, studies of the use of bowing parameters and playability take a limited range of bow velocity into account, covering only a small portion of the parameter range used in real performance. The 16th-note condition in this study allowed extending the analysis of the used bowing-parameter space to higher bow velocities up to 2 m/s.

In order to provide an overview of the various conditions with a wide range of bow velocity, an alternative graphical representation was chosen, as the Schelleng diagram is in principle only valid at a single value of bow velocity. In the alternative representation, F_B is plotted versus the v_B/β ratio with both axes logarithmically scaled. The maximum Schelleng limit can then be represented by a straight line with slope 1. The lower bow-force limit cannot be simply shown in this diagram for two reasons. Firstly, Schelleng's lower bow-force limit is not uniquely determined by the v_B/β ratio; at a fixed value of bow velocity the minimum Schelleng limit could be represented by a straight line with slope 2, the offset being dependent on v_B; on the other hand, at a fixed value of β, the minimum bow force is then represented by a straight line with slope 1, parallel to the maximum bow-force limit.[18] Secondly, the actual lower bow-force limit has not been reliably determined at high bow velocities, and it is unknown how the minimum bow force depends on bow velocity in a real bowed string. Measurements by Schoonderwaldt *et al.*[15] showed that the minimum bow force did not significantly depend on bow velocity in the range v_B=5–20 cm/s, and there were several indications that Schelleng's equation for minimum bow force did not provide an adequate explanation of the observations.

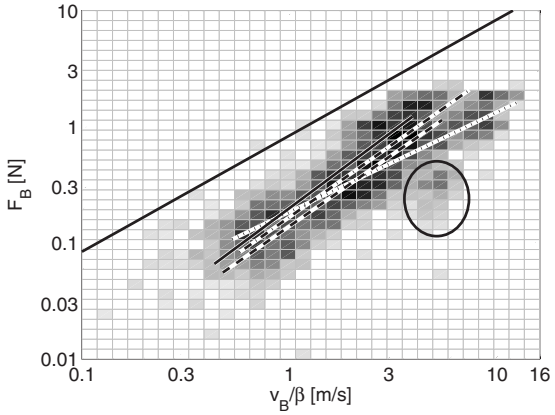


FIG. 3. Two-dimensional density distribution of F_B versus v_B/β (violin D string, including all players and all conditions). The upper bow-force limit (slope 1) is indicated by the upper solid line. The fitted lines represent the used v_B/β range and average trend of the four conditions: whole notes (solid line), half notes (dashed line), crescendo-diminuendo (dashed-dotted line), and 16th notes (dotted line). The circle indicates the performance of the 16th-note *pp* condition by one of the players, showing a deviating strategy.

Figure 3 shows the two-dimensional density distribution of the combined conditions (the long-note and 16th-note conditions at 3 dynamic levels, as well as crescendo-diminuendo conditions) in the alternative representation. There was a considerable overlap between the four included conditions (détaché whole notes, half notes, 16th notes, and crescendo-diminuendo). The used v_B/β range and average trend for each condition are indicated with fitted lines. There was a good agreement between the whole-note and half-note conditions. For the whole notes the fitted line was shifted to lower values of v_B/β and the slope was slightly steeper, indicating that the v_B/β ratio was generally smaller while the used bow forces were relatively high, especially at *f* level. In the crescendo-diminuendo condition the slope was similar to the half-note condition, but the range of v_B/β was extended to higher values, mainly due to the extended range in bow velocity. The slopes of the three long-note conditions (1.24–1.33) were all slightly steeper than the slope (1) of the upper bow-force limit.

In the 16th-note condition the behavior was different. The used range of v_B/β was considerably larger, but the bow forces observed at high values of v_B/β were generally not larger than in the other conditions and seemed to be limited to about 2.5 N. At *f* level, the maximum bow force of about 5–10 N was far from reached. As a result, the fitted slope (0.86) was much less steep, even smaller than that of the upper bow-force limit.

One particular area, indicated by a circle in Fig. 3,

clearly deviated from the diagonal. This area could be attributed to the performance of the 16th-note *pp* condition by one of the players, who used a relatively high bow velocity at small values bow-bridge distance. By listening to the recorded sound it could be concluded that even in this case Helmholtz motion was maintained most of the time. Many instances of multiple slipping could be detected at the bow changes, but that part of the tones was not considered in the above analysis. In contrast, on the G string, the same strategy led to multiple slipping throughout the entire condition, which could be clearly distinguished in the sound level (see Sec. IV A).

Figures 2 and 3 show that the three main bowing parameters as used by the players were clearly interrelated due to the influence of the constraints of the playable region in the Schelleng diagram. Generally, F_B was negatively correlated with β, and positively correlated with v_B; the correlation between v_B and β was somewhat weaker. The correlations per condition could be clearly visualized in the F_B versus v_B/β diagram. The R^2 values of the fitted slopes were rather high in all conditions, indicating a strong correlation between F_B and v_B/β (Pearson correlation $r \approx 0.9$).

C. Bowing strategies in setting dynamic level

In Helmholtz motion the amplitude of the string vibration is in principle determined by the v_B/β ratio, making bow velocity and bow-bridge distance the main control parameters of dynamic level. However, it is known that the *perceived* loudness is not only dependent on the peak-to-peak amplitude, but also on the distribution of energy in the spectrum, or rather, the distribution of energy across critical bands in the ear.[19] Since the amount of corner rounding—and thus the energy of the higher partials—is mainly influenced by bow force,[16,20] F_B also has a significant influence on the perceived dynamic level.

As F_B was above shown to be highly correlated with v_B/β, the following analysis will focus on the role of bow velocity and bow-bridge distance. Figure 4 shows the used ranges of these parameters, along with the resulting v_B/β ratio in the violin performances for the three note-length conditions and the combined crescendo-diminuendo conditions. In the whole-note and half-note conditions bow velocity was clearly constrained by the length of the bow, and the range in bow velocity across dynamic levels was small. The variation in v_B/β was clearly dominated by the contribution of β, especially in the whole-note condition.

In the whole-note condition most players used the full length of the bow, corresponding to a bow velocity of 15 cm/s, and reduced bow velocity only slightly to about 11 cm/s when playing *pp*. In the half-note condition, the players reduced the average bow velocity from 30 down to 18 cm/s with a decreasing dynamic level. These observations are in agreement with Askenfelt's findings[9] that players preferred bow velocities in the range 20–40 cm/s, and only occasionally brought down the bow velocity to 10 cm/s.

For the 16th-note conditions, the opposite was found. The used range of bow velocity was extensive, showing large differences between dynamic levels, and bow velocity was

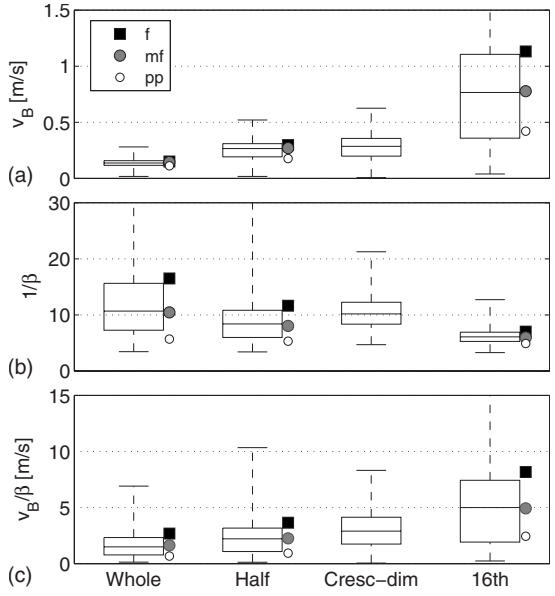


FIG. 4. Box plots of (a) v_B, (b) $1/\beta$, and (c) v_B/β for the three note-length conditions (all three dynamic levels included), as well as the four combined crescendo-diminuendo conditions (half notes). In panel (b) the reciprocal of β is shown for a better indication of its influence on dynamic level. The distributions include the collected data from all violin players and all four strings. The average values of the parameters per dynamic level are marked.

clearly dominant in setting the string amplitude. The used range of β was rather small and did not show large differences between dynamic levels.

The minimum bow-bridge distance in the 16th-note condition was significantly larger compared to the long-note conditions. There might be several reasons for this. Firstly, a high bow velocity in combination with a small bow-bridge distance entail the danger of partial slip phases due to the incompatibility of the string displacement with the finite width of the bow, which would result in audible noise.[21] A second possible explanation might be related to the bow change. In the long-note conditions at forte level the peak accelerations during the bow change were mainly in the range 10–20 $\mathrm{m/s^2}$, whereas in the 16th-note condition bow acceleration ranged from 20 to 50 $\mathrm{m/s^2}$. In the latter case, a larger bow-bridge distance provides more benign conditions for the creation of Helmholtz motion,[5] which might partly explain the strategy chosen by the players.

In the crescendo-diminuendo conditions the range of bow velocity used was extended to higher bow velocities compared to the half notes at fixed dynamic levels, indicating that bow velocity played a more prominent role in the variation of dynamic level.

D. Individual differences between players

As the number of players participating in this study was limited, it is important to consider the individual strategies. Figure 5 shows the average values of the main bowing parameters in the different conditions for the three violinists. Generally, the individual participants showed similar trends between conditions as shown above. Bow velocity was appreciably higher in the 16th-note condition for all players,

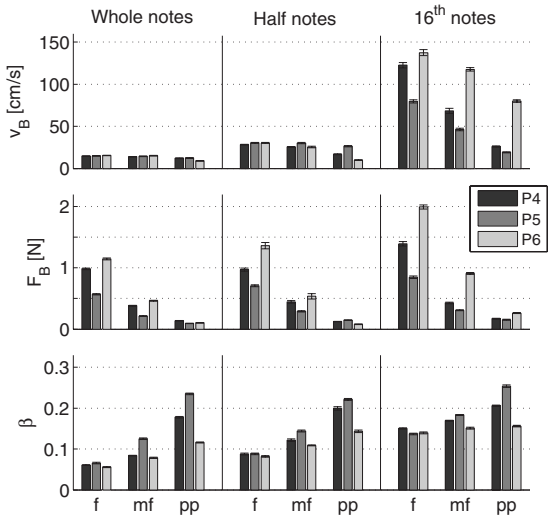


FIG. 5. Average bowing parameters per condition for the three violinists (P4, P5, and P6) across all strings. The error bars indicate 95% confidence intervals of the means.

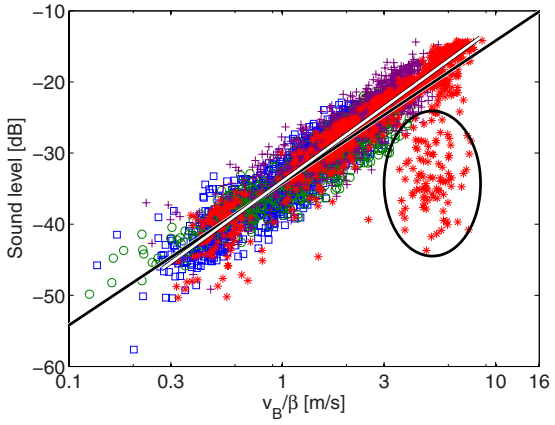


FIG. 6. (Color online) Sound level (dB) versus v_B/β (logarithmically scaled) for all conditions [whole notes (□), half notes (○), and 16th notes (∗) at *pp*, *mf*, and *f* levels, and crescendo-diminuendo half notes (+), in the online version also indicated by different colors] performed by the three violinists on the G string. The black line indicates the theoretical dependence with a slope of 20 dB/decade. The points within the circle clearly deviated from the diagonal, and could be attributed to prolonged episodes of multiple slipping in the performance of the 16th-note *pp* condition by one of the players. The white line indicates the fitted linear relation (slope: 21.7 dB/decade, $R^2=0.90$; the outliers indicated by the circle were discarded).

and the range of β was extended to smaller values in the long-note conditions. Bow force was only slightly higher in the 16th-note condition.

In the long-note conditions there were only minor differences in the use of bow velocity between the players; only in the half-note *pp* condition players P4 and P6 clearly reduced the bow velocity. However, in the 16th-note condition, the differences between players were substantial. Player P6 used relatively high bow velocities at all dynamic levels: 80 and 138 cm/s in the *pp* and *f* conditions, respectively, corresponding to a factor 1.7. In comparison, player P5 used 19 and 80 cm/s (a factor of 4.2). Player P4 showed the largest range in bow velocity, 26–123 cm/s (a factor of 4.7).

Player P6 used by far the highest bow forces in all conditions but one, especially at forte level (from around 1 N in whole notes to 2 N in 16th notes). She also used the largest range in force across dynamic levels (e.g. 0.25–2 N in 16th notes). Player P5 used relatively low bow forces, typically less than half the force values for P6, except in *pp*. Player P4 could be placed in between.

Regarding bow-bridge distance, there was a high degree of agreement between players in the *f* conditions. Interestingly, β was approximately doubled by all three players from about 1/14 (17 mm from the bridge) in the whole-note *f* condition to 1/7 (34 mm) in the 16th-note *f* condition. However, at *mf* and *pp* levels, there were substantial individual differences. Player P6 used only a limited range of β around 1/10 (25 mm). Player P5 used the largest range of β between conditions (1/15–1/4, corresponding to 16–63 mm) utilizing more the fingerboard area. Player P4 could again be placed in between.

IV. RELATION BETWEEN MAIN BOWING PARAMETERS AND SOUND FEATURES

A. Sound level

Figure 6 shows the relation between sound level and v_B/β for the combined data of all violinists (P4, P5, and P6) and the combined conditions (whole notes, half notes, 16th notes, and crescendo-diminuendo) on the violin G string. There was a good overlap between the conditions, and the combined clusters showed a strong linear relation ($R^2=0.90$, outliers discarded). The fitted slope (21.7) was close to the theoretical value of 20 dB/decade.

There were some notable outliers, as indicated by a circle. These could be attributed to the performance of the 16th-note *pp* condition by one of the players, which was dominated by multiple slipping. The proportionality of string amplitude to v_B/β is lost in that case, and the vibrating string does no longer reach its full amplitude, leading to a significant decrease in sound level.

Similar results were found for the higher strings. The R^2 values (0.82–0.90) were generally high, indicating that the sound level was highly correlated with v_B/β (in logarithmic scale). The fitted slopes (20.2–21.7) were slightly but significantly larger than the theoretical 20 dB/decade for all four strings. This might possibly be explained by the influence of bow force; the coordinated increase in F_B with v_B/β causes an increase in the energy of the higher partials, and the influence on the details of the wave shape might boost the rms value used for the calculation of sound level.

Table II shows the average sound levels per condition relative to the sound level of the whole-note *pp* condition. The value predicted by the v_B/β ratio is indicated between

TABLE II. Average sound level (in decibel) per note-duration condition and dynamic level on the violin G string (all three players included). The values predicted by the v_B/β ratio (based on the geometric average) are indicated between parentheses. All levels were calculated relative to the whole-note *pp* condition.

	Whole	Half	16th
f	11.0 (10.1)	13.9 (13.4)	20.3 (18.7)
mf	7.2 (7.1)	11.1 (11.0)	15.0 (14.2)
pp	0 (0)	1.9 (2.1)	4.2 (4.2)

TABLE III. Contributions in decibel of bow velocity and bow-bridge distance to the $pp-f$ range of predicted sound level (last column) per note-duration condition. The β-weights obtained by multiple regression with sound level as dependent variable and v_B and $1/\beta$ (logarithmically scaled) as independent variables are shown between parentheses. The values are based on the performances of all three violinists on the G string.

	v_B	$1/\beta$	Total
Whole notes	2.7 (0.30)	7.4 (0.81)	10.1
Half notes	5.7 (0.53)	5.6 (0.56)	11.3
16th notes	11.0 (0.86)	3.5 (0.15)	14.5
Crescendo-diminuendo	- (0.66)	- (0.41)	-

parentheses. The agreement between the predicted and observed values of sound level confirms that sound level could be well predicted by the combination of bow velocity and bow-bridge distance.

The table shows that sound level was highest in the 16th-note f condition. The sound level for a given dynamic level depended on note duration; for example, the average sound level in the 16th-note mf condition exceeded that in the whole-note f condition. The maximum difference of dynamic levels across note-duration conditions was about 20 dB. Within note-duration conditions the average ranges of sound level were between 11 and 16 dB. Earlier reported differences between dynamic levels are 9–10 dB between p and f and 14 dB between pp and ff.[9,22]

The sound level range per string spanning from 2.5% to 97.5% of the distribution (i.e., containing 95% of the measured values) was about 30 dB (from the G to the E string: 28, 26, 28, and 34 dB). Taking all four strings into account, the 95% sound level range was about 31 dB, corresponding to the maximum dynamic range of sound with musically acceptable tone quality reported by Askenfelt.[9] The 99% range was 37 dB, which is in close agreement with earlier reported values,[23,24] as well as the total dynamic span found by Askenfelt[9] when not paying full respect to the tone quality.

In Table III the separate contributions (in decibel) of bow velocity and bow-bridge distance to the $pp-f$ range of predicted average sound level are shown for each note-length condition. The β-weights indicating the relative weight of the bowing parameters in determining sound level are shown between parentheses. The values clearly reflect the varying role of bow velocity and bow-bridge distance in setting the dynamic level for the different note-duration conditions, as shown in Sec. III C. In the whole-note condition the contribution of bow-bridge distance was dominant with 73%, whereas in the 16th-note condition bow velocity dominated with 76%. In the half-note condition bow velocity and bow-bridge distance contributed rather equally to the total sound level.

For the crescendo-diminuendo conditions the $pp-f$ sound level range could not be specified, as these conditions did not comprise discrete dynamic levels. However, the β-weights obtained by multiple regression indicate that bow velocity was the dominant parameter in varying sound level.

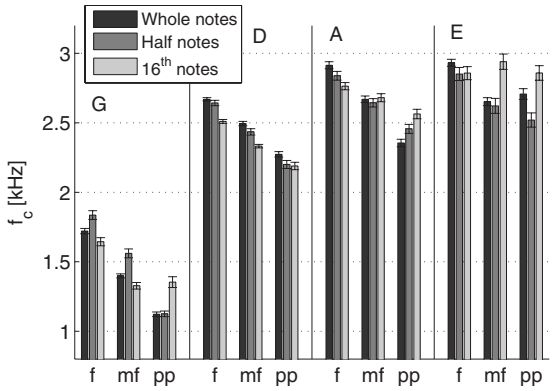


FIG. 7. Spectral centroid per string per condition averaged across all three violin players.

B. Spectral centroid

In Fig. 7 the spectral centroid per string per condition is shown averaged across the performances of all three violin players. As expected, the spectral centroid increased from the lowest to the highest string, however, not in proportion to fundamental frequency. The largest jump was observed between the G and the D string. The latter indicates that the effect of corner rounding was stronger on the G string, possibly due to its higher stiffness.

Spectral centroid showed a clear dependence on the performed task. Spectral centroid increased from pp to f, which could be expected from the increase in bow force. The contrast of spectral centroid was larger in the long-note conditions compared to the 16th-note condition.

The values of spectral centroid were in reasonable agreement with those found in an earlier experiment using a bowing machine.[16] The values of spectral centroid in that study for the corresponding combinations of bowing parameters in the whole-note condition were about 2.1, 1.6, and 1.3 kHz for f, mf, and pp, respectively, compared to 2.7, 2.5, and 2.3 kHz in the current study (D string). The values of spectral centroid in the bowing machine study were somewhat lower and showed a larger range. This might be due to the different way spectral centroid was calculated in that study, making direct comparison difficult.[25]

The dependence of spectral centroid f_c on the main bowing parameters was, in analogy with the bowing machine study,[16] analyzed using multiple linear regression. To optimize the regression model, a curvilinear transformation was applied to bow force. In the earlier study it was suggested that the dependence of f_c on F_B could be described by a power relation $f_c \propto F_B^{\alpha}$, with α between 0 and 1.[16] Fitting this relation to the data gave a value of $\alpha \approx 0.2$.

The regression yielded similar results as found in the bowing machine study.[16] On the G string the R^2 value was 0.71, indicating that the model accounted for a reasonable amount of the variance. The β-weights of F_B, v_B, and β (see Table IV) indicated that bow force was the dominating factor in controlling the spectral centroid, followed by bow velocity and bow-bridge distance, respectively. The negative sign of the coefficient of v_B indicates that spectral centroid decreased with increasing bow velocity. Likewise, the positive

TABLE IV. β-weights and R^2 values of regression model with spectral centroid as a dependent variable and bowing parameters as independent variables. The last column indicates the measured range of spectral centroid in kilohertz (from 2.5% to 97.5% of the distribution).

	F_B^α	v_B	β	R^2	95% range (kHz)
G	1.16	−0.51	0.18	0.71	0.9–2.2
D	0.84	−0.39	−0.05	0.60	1.9–2.9
A	0.69	−0.24	0.11	0.30	2.0–3.3
E	0.28	−0.11	0.11	0.03	2.0–3.5

sign of the coefficient of β indicates that spectral centroid increased with increasing bow-bridge distance.

The partial contributions of the independent variables F_B^α, v_B, and β to the spectral centroid are shown in Fig. 8. The range of variation of spectral centroid accounted for by bow force was about 2 kHz, followed by bow velocity (about 1 kHz) and bow-bridge distance (less than 0.5 kHz). The panels show clear linear relationships (i.e., *after* curvilinear transformation of F_B), indicating that spectral centroid could be well explained by the combination of bowing parameters. It should be noted that the contribution of bow force is in practice opposed by the contribution of bow velocity and bow-bridge distance, given the strong correlation between those parameters in playing. The resulting width of the range of spectral centroid per string was typically 1.0–1.5 kHz (see Table IV).

Table IV shows the outcome of the regression model for all four strings. The β-weights showed a similar trend across the strings. The R^2 value was highest for the G string; for the D and the A string the model explained 60% and 30% of the variance, respectively, and for the E string the regression model accounted for merely 3%. A possible reason might be that the players used vibrato, which especially on the higher strings resulted in sympathetic resonances (the notes played were one octave above the adjacent lower strings), causing large fluctuations of the spectral centroid. Also other factors might have played a role. The damping caused by the finger stopping the string is likely to have an influence on the spectrum as it plays an important role in the process of corner rounding. Variations in finger pressure due to vibrato might therefore cause additional fluctuations in spectral centroid without a direct relation with the bowing parameters.[26] On the E string, which has a low characteristic impedance and a low internal damping, fluctuations in damping might have a larger influence on the spectrum compared to the lower strings.

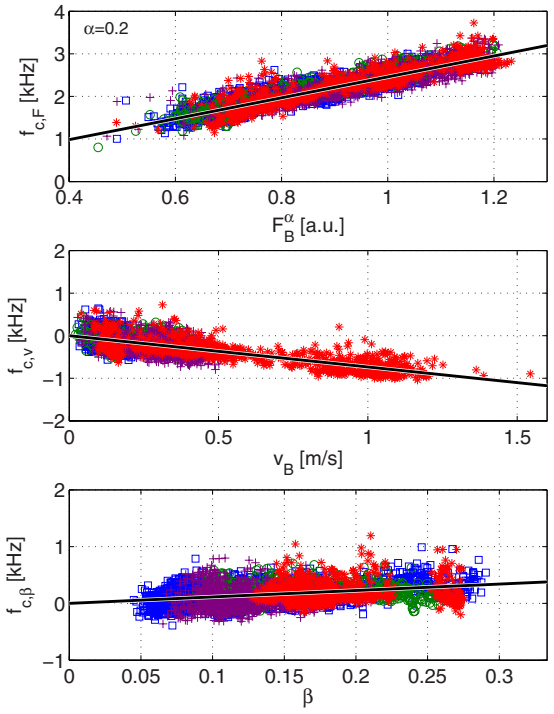


FIG. 8. (Color online) Partial contributions of F_B^α (with α=0.2), v_B, and β to the spectral centroid, including four conditions [whole notes (□), half notes (○), and 16th notes (∗) at *pp*, *mf*, and *f* levels, and crescendo-diminuendo half notes (+), in the online version also indicated by different colors]. The solid lines indicate the linear relations in the regression model. Violin, G string, including all conditions and players [multiple slipping (16th-note *pp* performance by one of the players) discarded].

V. ADAPTATION OF THE BOWING PARAMETERS TO STRING AND INSTRUMENT

A. Influence of string properties

Figure 9 shows the averages of the main bowing parameters per string on the violin. The bow force (middle panel) used on the G string was generally higher compared to the higher strings. This is in accordance with expectations, as the higher characteristic impedance, along with the higher inter-

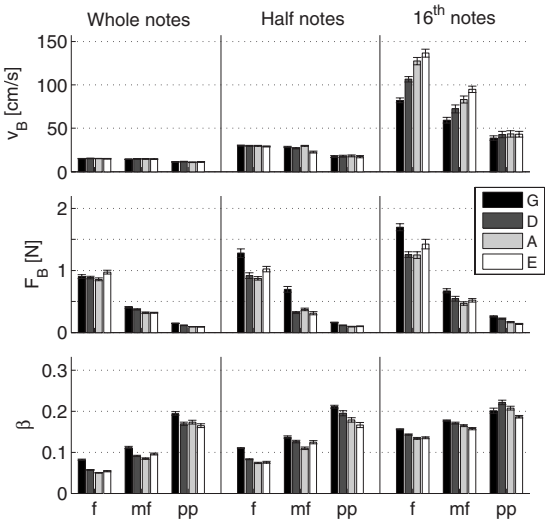


FIG. 9. Average bowing parameters per condition and string for the violin across all players.

TABLE V. Average values of v_B/β per string of the 16th-note f performances and the resulting peak amplitude $\hat{\eta}_{max}$ in the middle of the string and the pitch rise due to the increase in effective string tension. The latter values are based on empirically determined Young's moduli of the used strings.

String	v_B/β (m/s)	$\hat{\eta}_{max}$ (mm)	Pitch rise (cent)
G	5.2	2.5	8
D	7.5	2.4	6
A	9.5	2.0	4
E	10.5	1.5	5

nal damping of the G string, leads to higher upper and lower bow-force limits compared to the other strings. In the f conditions the force used on the E string was also consistently higher compared to the two middle strings (D and A). A possible explanation for this somewhat surprising behavior might be related to the geometry of the instrument. On the middle strings the bow force must be applied more carefully to prevent the bow from touching the neighboring strings. On the outer strings (G and E) the bow inclination is less constrained, giving the player more freedom to apply higher bow forces.

Furthermore, the average bow-bridge distance (lower panel) was slightly, but consistently, larger on the G string compared to the other strings, especially in the two long-note conditions. This might be another adaptation of the players to the higher bow-force limits for the G string. By increasing the bow-bridge distance, the increase in bow force required on the G string compared to the three higher strings will be smaller. A larger bow-bridge distance also facilitates bow changes in heavier strings (a higher Z_0).

Regarding bow velocity (upper panel), there were no noteworthy differences between different strings in the long-note conditions. However, the 16th-note f and mf conditions showed a clear increase in bow velocity from lower to higher strings. The amplitude of string vibration, which is proportional to the fundamental period, might offer a possible explanation. A large string amplitude gives rise to an increase in pitch due to an increase in effective string tension, an effect which players might want to avoid. This effect is especially noticeable on low strings, in particular, when they have a steel core.

The average values of v_B/β and the resulting amplitude (maximum displacement amplitude at the middle of the string, given by $\hat{\eta}_{max}=v_B T_1/8\beta$)[20] and pitch rise for each string are shown in Table V. The pitch rise under influence of the bowing parameters was estimated for the used strings via the procedure described by Schoonderwaldt (see Appendix in Ref. 16). It can be seen that the maximum string amplitude increased from the higher to the lower strings (E: 1.5 mm; G: 2.5 mm), despite the fact that the players halved the v_B/β ratio. If the same bow velocity as for the E string would have been used on the G string, the maximum amplitude would have been 5 mm, which would have resulted in a pitch rise of 33 cent instead of 8 cent. Interestingly, the calculated pitch rise is small and differs only a factor 2 across all 4 strings (4–8 cent).

It is further interesting to note that the peak transverse bridge force[27] was 1.8–1.9 N for all strings, indicating that the driving force on the violin bridge was almost equal for all four strings, despite the fact the Z_0 was a twice as high for the G string as for the E string. An equal driving force might have been another reason for players to adapt bow velocity to the specific properties of the string (characteristic impedance in this case). The motivation for an equal driving force is, however, not clear-cut. Obtaining a similar sound level and timbre on all strings is not warranted by an equal bridge force, as the interfacing bridge has rather different impedance transformation ratios for the G and E strings. In any case, the observation is an interesting aspect on the choice of strings (Z_0, Young's modulus), which matches playing style as well as instrument properties.

An alternative explanation for the reduced bow velocity on lower strings can be found in the attack. It is more difficult to obtain short Helmholtz transients during attacks or bow changes in strings with a higher characteristic impedance.[5] The players might therefore have used lower accelerations, resulting in lower peak velocities in fast notes characterized by sine-like velocity patterns (see Fig. 1).

B. Comparison between violin and viola

The main difference between violin and viola performances seems to lie in the use of bow force. The bow force in viola performance was generally higher than in violin performance, on average 0.3 N higher at f level. The viola C string was played with the highest bow force in all conditions, up to about 1 N higher than the violin G string.

The bow velocity was used in a very similar manner on both instruments. In the long-note conditions velocity was constrained to the length of the bow in whole-bow strokes, and in the 16th-note condition bow velocity was reduced from higher to lower strings. Similar to the violin performances, there were considerable individual differences in the use of bow velocity in the 16th-note performances. Interestingly, players P4 and P5, who performed on both instruments, used *higher* bow velocities on the viola compared to the violin, especially at softer dynamic levels. This might be related to a different approach in sound production. A similar observation was made by Chen,[28] when analyzing scales played at p and f levels recorded in the same experiment. However, given the small amount of participants in the experiment, no final conclusions can be drawn.

There were no noteworthy overall differences in the use of β between violin and viola performances. This implies that the *absolute* bow-bridge distance was adapted to the length of the string.

VI. DISCUSSION AND CONCLUSIONS

The analyses presented in this study provided a detailed description of the use of bowing parameters in the steady part of détaché notes, giving a deepened insight in the interaction between the player and the instrument. The distributions and averages represented many notes with a total effective playing time of about 15 min, allowing for a detailed

analysis of the different factors influencing the choice of bowing parameters, as well as individual differences between players.

The results of this study were generally in agreement with earlier findings by Askenfelt.[8,9] However, in the current study, the display of distributions of bowing parameters, rather than selected points or averages, provides a more comprehensive overview of the use of the available bowing-parameter space. Similar typical ranges of bowing parameters were found in violin performance: v_B ranged from about 5 cm/s to 2 m/s, F_B from slightly less than 0.1 N to about 2–2.5 N (in extreme cases, not included in the current analyses, peak bow forces up to 4 N were observed), and β from about 1/22 to 1/4, corresponding to 11–63 mm on the stopped string (cf. 15–84 mm on an open string). In the current experiment, it seems that the bow forces observed at soft dynamics were somewhat lower than reported by Askenfelt,[9] who found that bow forces below 0.5 N were rarely used, and that the lowest bow force easily accessible (at the tip) was about 0.15 N. In the current study bow forces around 0.1 N were normally observed at *pp* level, and even at *mf* level bow force was on average found to be slightly below 0.5 N.

It was shown that the strategy for the production of dynamic differences involved a trade-off between bow velocity and bow-bridge distance, confirming earlier findings by Askenfelt.[9] In the current study the range of note durations was further expanded, leading to higher contrasts in the observed control strategies, as well as larger differences between the extreme values of the bowing parameters. For the longest notes (4 s) dynamic level was almost entirely determined by bow-bridge distance, whereas for the 16th notes (0.2 s) bow velocity was the dominating parameter. Bow-bridge distance showed much less variation between dynamic levels, mainly because small values of bow-bridge distance were avoided at high bow velocities.

There were several indications of that players adapted the bowing parameters to the physical properties of the strings and the instrument. Bow force was clearly highest on the lowest strings, especially on the C string on the viola. This indicates that players are sensitive to the differences in bow-force limits across strings, constantly optimizing their performance. On the viola higher bow forces were used in general, in accordance with the higher characteristic impedances of the strings. Also bow velocity in détaché 16th notes performed at *f* level was adapted to the string played: Bow velocity was decreased from higher to lower strings, possibly to limit the amplitude of vibration and to facilitate the bow changes.

The distributions of bow force and bow-bridge distance in the Schelleng diagram (Fig. 2) showed that the (empirically determined) bow-force limits were generally well respected. When changing the dynamic level the players moved diagonally through the Schelleng diagram, following the contours of the bow-force limits. Bow force was also observed to increase with bow velocity. However, at extreme bow velocities in loud 16th notes, the full range of bow force was not utilized. The bowing parameters were highly interrelated, as indicated by the high correlations observed for F_B versus v_B/β across dynamic level.

Sound level and spectral centroid showed a clear dependence on the main bowing parameters. As expected, the v_B/β ratio was found to play a major role in setting dynamic level. However, the coordinated increase in bow force with v_B/β also led to an increase in spectral centroid with increasing dynamic level. As the perceived loudness is not only dependent on amplitude, but also on the energy of the higher partials, this will reinforce the perceived contrast between dynamic levels. The role of spectral centroid in perceived dynamic level might be more important in long notes, in which the contrast of spectral centroid was found to be larger than in fast détaché notes (see Fig. 7).

Concerning spectral centroid, the analyses confirmed earlier observations by Guettler *et al.*[29] and Schoonderwaldt.[16] Similar results were also obtained by Demoucron[13] in an extensive study of the influence of bowing parameters on the sound produced by a virtual violin (physical model). It was found that bow force was by far the most dominant control parameter. The spectral centroid increased with increasing bow force and decreased with increasing bow velocity. It was further confirmed that spectral centroid increased slightly with increasing bow-bridge distance. The latter result is rather counterintuitive, but can be explained by the high correlation between bow-bridge distance and bow force: Bow force is mostly decreased when bow-bridge distance is increased, leading to a net decrease in spectral centroid.

It can be concluded that our method for measuring bowing gestures allows for a detailed analysis of bowing strategies and subtle aspects of control, taking most relevant parameters into account. It is hoped that the current study will contribute to a deepened understanding of tone production in string instrument performance, as well as of the interaction between the player and the instrument. Only a small portion of all the subtleties, which can be readily observed in the data, could be accounted for in the current study. Preliminary analyses of some of these aspects have been presented in Ref. 30 including the use and possible control functions of the bow angles *tilt* and *skewness*. Follow-up studies will be needed to focus on more advanced aspects of bow control, such as attacks, bow changes, and complex note patterns involving string crossings.

ACKNOWLEDGMENTS

This work was partly supported by the Swedish Science Foundation (Contract No. 621-2001-2537) and the Natural Sciences and Research Council of Canada (NSERC-SRO). The experiment was performed in collaboration with Lambert Chen, who selected the musical fragments, and greatly contributed to the design of the tasks and the recruitment of the participants for the experiment. Part of the results were incorporated in his D.Mus. thesis. Many thanks go to Marcelo M. Wanderley and Anders Askenfelt for their inspiring supervision, and to the players for their enthusiastic participation in the experiment.

[1]H. Winold, E. Thelen, and B. D. Ulrich, "Coordination and control in the bow arm movements of highly skilled cellists," Ecological Psychol. **6**, 1–31 (1994).
[2]J. Konczak, H. vander Velden, and L. Jaeger, "Learning to play the violin: Motor control by freezing, not freeing degrees of freedom," J. Motor Behav. **41**, 243–252 (2009).
[3]C. V. Raman, "Experiments with mechanically-played violins," Proc. Indian Assoc. Cultivation Sci. **6**, 19–36 (1920–1921) [Reprinted in *Musical Acoustics, Part I*, C. Hutchins (ed.), Dowden, Hutchinson & Ross, Inc., Stroudsburg, PA].
[4]J. C. Schelleng, "The bowed string and the player," J. Acoust. Soc. Am. **53**, 26–41 (1973).
[5]K. Guettler, "On the creation of the Helmholtz motion in bowed strings," Acta Acust. Acust. **88**, 970–985 (2002).
[6]P. M. Galluzzo, "On the playability of stringed instruments," Ph.D. thesis, Trinity College, University of Cambridge, Cambridge, UK (2003).
[7]S. Serafin, "The sound of friction: Real-time models, playability and musical applications," Ph.D. thesis, Stanford University, Stanford, CA (2004).
[8]A. Askenfelt, "Measurement of bow motion and bow force in violin playing," J. Acoust. Soc. Am. **80**, 1007–1015 (1986).
[9]A. Askenfelt, "Measurement of the bowing parameters in violin playing. II: Bow-bridge distance, dynamic range, and limits of bow force," J. Acoust. Soc. Am. **86**, 503–516 (1989).
[10]D. Young, "A methodology for investigation of bowed string performance through measurement of violin bowing technique," Ph.D. thesis, Massachusetts Institute of Technology, Boston, MA (2007).
[11]A. Perez, J. Bonada, E. Maestre, E. Guaus, and M. Blaauw, "Combining performance action with spectral models for violin sound transformation," in Proceedings of the 19th International Congress on Acoustics (ICA07), Madrid, Spain (2007).
[12]N. Rasamimanana, "Geste instrumental du violoniste en situation de jeu: Analyse et modélisation (Violin player instrumental gesture: Analysis and modelling)," Ph.D. thesis, Université Pierre et Marie Curie (UPMC), Paris VI, France (2008).
[13]M. Demoucron, "On the control of virtual violins: Physical modelling and control of bowed string instruments," Ph.D. thesis, Paris & Royal Institute of Technology (KTH), Université Pierre et Marie Curie (UPMC), Stockholm, Sweden (2008).
[14]E. Schoonderwaldt and M. Demoucron, "Extraction of bowing parameters from violin performance combining motion capture and sensors," J. Acoust. Soc. Am. **126**, 2695–2708 (2009).
[15]E. Schoonderwaldt, K. Guettler, and A. Askenfelt, "An empirical investigation of bow-force limits in the Schelleng diagram," Acta Acust. Acust. **94**, 604–622 (2008).
[16]E. Schoonderwaldt, "The violinist's sound palette: Spectral centroid, pitch flattening and anomalous low frequencies," Acta. Acust. Acust. **95**, 901–914 (2009).
[17]The limits of the "gray zones" indicated by the hatched areas are based on the fitted Schelleng limits on an open D string on a monochord and a stopped D string on a violin (the same violin as used in the current study, but a different string, stopped at pitch E♭4). The lower bow-force limit was considered to be independent of bow velocity, as found by Schoonderwaldt *et al.* (Ref. 15).
[18]A comparable two-dimensional representation of the bowing-parameter space was proposed by Askenfelt (Ref. 9) with v_B/β^2 on the abscissa, instead of v_B/β. This representation is more favorable for visualization of the minimum bow force limit, under the assumption that Schelleng's equation for minimum bow force holds.
[19]B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 5th ed. (Elsevier Academic, London, 2004).
[20]L. Cremer, *The Physics of the Violin* (MIT, Cambridge, MA, 1984).
[21]M. E. McIntyre, R. T. Schumacher, and J. Woodhouse, "Aperiodicity in bowed-string motion: On the differential-slipping mechanism," Acustica **50**, 294–295 (1982).
[22]M. Clark and D. Luce, "Intensities of orchestral instrument scales played at described dynamic markings," J. Audio Eng. Soc. **13**, 151–157 (1965).
[23]J. S. Bradley, "Effect of bow force and speed on violin response," J. Acoust. Soc. Am. **60**, 274–275 (1976).
[24]J. Meyer, "The sound of the orchestra," J. Audio Eng. Soc. **41**, 203–213 (1993).
[25]In the bowing machine study (Ref. 16) spectral centroid was calculated from the string velocity signal under the bow, which required a modification of the spectrum. In both the current study and the bowing machine study a limiting frequency of 10 kHz was used in the calculation of the spectral centroid.
[26]Some interesting observations on finger pressure have been made by H. Kinoshita, S. Obata, H. Nakahara, S. Furuya, and T. Aoki ["Fingering force during violin playing: Tempo, loudness and finger effects in single sound production," poster presented at Neuroscience and Music III, Montreal, QC, Canada (2008)].
[27]The peak transverse bridge force was calculated using Eq. (3.23) in Ref. 20.
[28]L. J.-Y. Chen, "A modern study of viola playing," D.mus. thesis, Schulich School of Music, McGill University, Montreal, Canada (2008).
[29]K. Guettler, E. Schoonderwaldt, and A. Askenfelt, "Bow speed or position—Which one influences spectrum the most?," in Proceedings of the Stockholm Music Acoustics Conference (SMAC03), Stockholm, Sweden (2003), pp. 67–70.
[30]E. Schoonderwaldt, "Mechanics and acoustics of violin bowing: Freedom, constraints and control in performance," Ph.D. thesis, KTH—School of Computer Science and Communication, Stockholm, Sweden (2009).

Vocal cues to identity and relatedness in giant pandas (*Ailuropoda melanoleuca*)

Benjamin D. Charlton[a)]
Zoo Atlanta, 800 Cherokee Avenue S.E., Atlanta, Georgia 30315-1440

Zhang Zhihe
Chengdu Research Base of Giant Panda Breeding, 26 Panda Road, Northern Suburb, Chengdu, Sichuan Province 610081, People's Republic of China

Rebecca J. Snyder
Zoo Atlanta, 800 Cherokee Avenue S.E., Atlanta, Georgia 30315-1440

(Received 15 December 2008; revised 13 August 2009; accepted 15 August 2009)

A range of acoustic characteristics typically carry information on individual identity in mammalian calls. In addition, physical similarities in vocal production anatomy among closely related individuals may result in similarities in the acoustic structure of vocalizations. Here, acoustic analyses based on source-filter theory were used to determine whether giant panda bleats are individually distinctive, to investigate the relative importance of different source-(larynx) and filter-(vocal tract) related acoustic features for coding individuality, and to test whether closely related individuals have similarities in call structure. The results revealed that giant panda bleats are highly individualized and indicate that source-related features, in particular, mean fundamental frequency, amplitude variation per second, and the mean extent of each amplitude modulation, contribute the most to vocal identity. In addition, although individual pairwise relatedness was not correlated with overall acoustic similarity, it was highly correlated with amplitude modulation rate and fundamental frequency range, suggesting that these acoustic features are heritable components of giant panda bleats that could be used as a measure of genetic relatedness. The ecological relevance of acoustically signaling information on caller identity and the potential practical implications for acoustic monitoring of population levels in this endangered species are discussed. [DOI: 10.1121/1.3224720]

PACS number(s): 43.80.Ev [MJO] Pages: 2721–2732

I. INTRODUCTION

Recent studies of mammal vocal communication have used the source-filter theory of human speech production (Fant, 1960) as a causal framework with which to understand vocal production and acoustic structure (Fitch and Hauser, 1995; Fitch, 1997; Owren *et al.*, 1997; Reby and McComb, 2003; Rendall *et al.*, 2005; Riede *et al.*, 2005; Harris *et al.*, 2006; Sanvito *et al.*, 2007). This theory states that mammal vocal signals are generated by the conversion of airflow from the lungs to acoustic energy by the larynx, the source, which is subsequently filtered by the vocal tract. This vocal tract filtering selectively amplifies certain frequencies, which are called vocal tract resonances or formants. Because inter-individual differences in laryngeal and vocal tract morphology are likely, both source and filter-related acoustic characteristics have the potential to yield information on a given caller's identity/phenotype. In this study, we investigate the relative importance of different source- and filter-related acoustic features of giant panda (*Ailuropoda melanoleuca*) bleats for coding individuality and relatedness.

Evidence that vocal signals contain sufficient information for vocal recognition (Reby *et al.*, 1998, 1999; Semple, 2001; McCowan and Hooper, 2002; Charrier *et al.*, 2003; McComb *et al.*, 2003; Rendall, 2003; Yin and McCowan, 2004; Blumstein and Munos, 2005; Soltis *et al.*, 2005; Reby *et al.*, 2006; Vannoni and McElligott, 2006) and that conspecifics can discriminate between individuals based on their vocalizations (Rendall *et al.*, 1996; Hare, 1998; Sayigh *et al.*, 1999; Charrier *et al.*, 2001; Reby *et al.*, 2001; Frommolt *et al.*, 2003) is documented in numerous mammal species. Indeed, we would expect signalers to be individually distinctive and for receivers to perceive discriminatory cues where it is adaptive for them to do so. For instance, in animals with social signaling systems, such as elephants (McComb *et al.*, 2000) and seals (Insley, 2000), selection to maximize differences between individual's vocalizations and for receivers to differentiate among callers would be expected. However, vocal recognition of unfamiliar and familiar individuals, and the ability to signal individual identity, may also be important in inter- and intra-sexual contexts. For example, males may use acoustic cues to identity to avoid unnecessary contests with known rivals (Tripovich *et al.*, 2008) and even to identify the sexual calls of specific females (Semple, 2001). In addition, females may become familiar with the vocalizations of certain males and preferentially mate with individuals that can afford higher energy courtship displays (East and Hofer, 1991; Zimmerman and Lerch, 1993; Reby *et al.*, 1998; McElligott *et al.*, 1999).

[a)]Author to whom correspondence should be addressed. Electronic mail: bcharlton@zooatlanta.org

Furthermore, because individual differences in vocal production anatomy affect the acoustic structure of mammal vocal signals (Fitch and Hauser, 1995, 2002), closely related individuals may have similar vocal characteristics, simply due to physical similarities in vocal production anatomy. However, distinguishing between the effects of vocal learning and genetic similarity on vocal characteristics is difficult (Janik and Slater, 1997; Egnor and Hauser, 2004). Indeed, although Zimmerman and Hafen (2001) showed that acoustic differences were correlated with genetic distances across three different colonies of Malagasy lemurs, studies explicitly comparing the known genetic relatedness of individuals with the acoustic structure of their vocalizations are rare in mammals (but see Crockford *et al.*, 2004).

The giant panda is particularly vocal during the breeding season (Kleiman *et al.*, 1979; Peters, 1982; Kleiman, 1983; Peters, 1985; Lindburg *et al.*, 2001). While the importance of olfaction in this species sexual communication is well documented (for a review see Swaisgood, 2004) until now, work on giant panda vocal communication had not progressed beyond outlining the vocal repertoire, giving basic data on acoustic structure and ascribing broad functional categories to specific vocalizations (Peters, 1982; Kleiman, 1983; Peters, 1985). The primary vocalization of the giant panda is a harmonically rich frequency and amplitude modulated "bleat" that encodes information on the caller's sex, age, and body size (Charlton *et al.*, in press). However, given its relative acoustic complexity, it also appears to be particularly well suited for coding individual identity (see Fig. 1).

Indeed, narrow-band frequency and amplitude modulated calls, such as the giant panda bleat, are generally thought to be the easiest to distinguish because they contain more information potentially available to receivers (Wiley and Richards, 1978). Moreover, the dense harmonic structure of giant panda bleats should highlight the formants, increasing their salience to receivers and making these calls well suited for identity cueing [as discussed by Owren and Rendall (1997, 2001)]. Furthermore, although giant pandas are not social animals, where selection for vocal individuality may be strong, they do occupy overlapping ranges and males roam widely (Schaller *et al.*, 1985). Consequently, because individuals will interact with others in adjacent territories, and especially during the breeding season (Schaller *et al.*, 1985), they may have the opportunity to familiarize themselves with the bleats of other conspecifics and acoustically identify individuals that they have previously encountered.

The goal of the current study was to determine whether giant panda bleats are individually distinctive vocalizations and to investigate the relative importance of different source- and filter-related acoustic features of giant panda bleats for coding individuality. We then go on to investigate whether acoustic similarity could represent a measure of genetic relatedness in giant pandas, and discuss the ecological relevance of acoustically signaling information on caller identity and the practical implications for non-invasively estimating population levels in this highly endangered species.

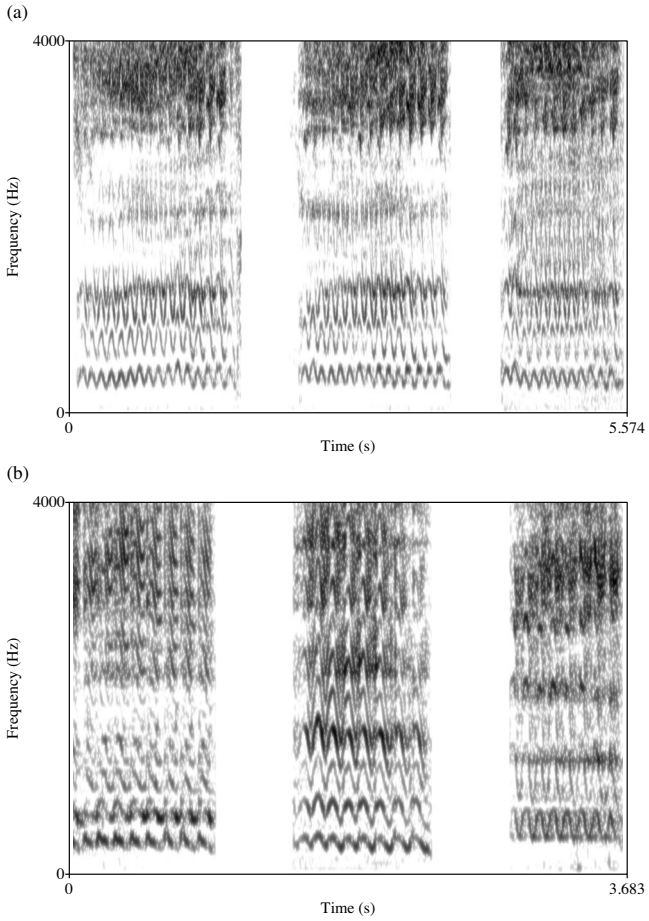


FIG. 1. Spectrograms of bleats from four different individuals. The top three spectrograms (a) show bleats from the same female and illustrate the low degree of intra-individual variability in bleat acoustic structure. The lower three spectrograms (b) show bleats from three different males and illustrate the high level of inter-individual variability in bleat acoustic structure (spectrogram settings: FFT method, window length: 0.03 s; time steps: 250; frequency steps: 1000; Gaussian window shape; and dynamic range: 50 dB).

II. MATERIALS AND METHODS

A. Study site and subjects

Bleats were recorded from 14 adult giant pandas at Chengdu Research Base of Giant Panda Breeding, Chengdu, China. In addition, bleats from four adult giant pandas resident in the United States at Zoo Atlanta, Memphis Zoo, and National Zoological Park, WA were recorded. This made a total of 18 subjects (9 males and 9 females) with ages ranging from 6 to 21 years (mean = 10.9 ± 3.62) (the sex, age, and sample sizes for each individual in the analysis are given in Table I). To minimize the effect of individual variability in arousal levels, bleats were recorded on at least two different days for each subject.

B. Recordings

The recordings were made using an Audio-Technica AT835b microphone and a TASCAM HDP2 portable solid-state digital recorder (sampling rate: 48 kHz, amplitude resolution: 16 bits) at distances ranging from 5 to 20 m. The recordings were transferred to an Apple Macintosh Macbook computer, normalized to 100% peak amplitude, and saved as

TABLE I. Age and sex of each subject in the analysis. N=number of bleats from each subject in the analysis.

Subject	N	Sex	Age (years)
Bing bing	10	Female	21
Cheng gong	29	Female	7
Cheng ji	9	Female	7
Da shuang	13	Female	10
Li li	10	Female	15
Lun lun	14	Female	10
Qi yuan	8	Female	8
Qi zhen	13	Female	8
Shu lan	18	Female	13
Bing dian	7	Male	7
Kebi	21	Male	15
Le le	13	Male	8
Lin lin	9	Male	10
Ping ping	14	Male	16
Tian tian	8	Male	10
Xiong bang	19	Male	6
Xiao shuang	15	Male	10
Yang yang	15	Male	10
Total	245 (139 male bleats and 106 female bleats)		

AIFF files (48 kHz sampling rate and 16 bit amplitude resolution). The overall spectral structure of each bleat was initially investigated using narrow-band spectrograms (see Fig. 1: fast Fourier transform (FFT) method; window length 0.03 s; time steps=250; frequency steps=1000; Gaussian window shape; and dynamic range=50 dB) and recordings with high levels of background noise were discarded. This left a total of 245 bleats from 18 individuals for the analysis (see Table I for the number of bleats recorded from each subject).

C. Acoustic analyses

Source- and filter-related acoustic features were extracted and measured using custom built programs in PRAAT 5.0.29 DSP package (Boersma and Weenink, 2005) that automatically logged these variables in an output file. PRAAT commands are included in parentheses.

1. Source-related measures

To characterize the source we measured a number of features from the fundamental frequency (F_0) contour (see Fig. 2). A cross-correlation algorithm [To Pitch (cc) command] was used to produce time-varying numerical representations of the F_0 contour for each bleat. The time step in the analysis was 0.01, and a five-point average smoothing filter was used to remove any rapid variations caused by analysis imprecision. To limit the possibility of "octave jumps," the minimum and maximum values for F_0 were set according to the F_0 contour as observed on the spectrogram. From the F_0 contour a number of parameters were extracted: call duration in seconds (duration); mean, minimum, and maximum F_0 across a call in hertz (mean, minimum, and maximum F_0, respectively); the F_0 range in hertz (F_0 range); the cumulative variation in the F_0 contour in hertz divided by call duration to give the mean variation per second (F_0 var); the number of complete cycles of F_0 modulation per second frequency modulation (FM rate); and the mean peak-to-peak variation of each F_0 modulation in hertz (FM extent) (see Fig. 2). More subtle F_0 variations were quantified using a measure of cycle-to-cycle frequency (or period) variation termed jitter (Titze, 1994). In this analysis jitter was calculated as the mean absolute difference between consecutive frequencies divided by the mean F_0 of each bleat and expressed as a percentage [Jitter (local) command].

2. Intensity measures

The intensity contour of each bleat was also extracted (To intensity command) to measure the cumulative variation in amplitude divided by call duration to give the mean variation per second in decibel (Ampvar), the mean peak-to-peak variation of each amplitude modulation in dB amplitude modulation (AM extent), and the number of complete cycles of amplitude modulation per second (AM rate) (see Fig. 2).

3. Filter-related measures

To characterize the filter the frequency values (in hertz) of the first six formants were measured using linear predictive coding [To Formants (Burg) command] and the following analysis parameters: time step: 0.01 s; window analysis: 0.20 s; maximum formant value: 3800–4000 Hz; maximum number of formants: 5–6; and pre-emphasis: 50 Hz. To check if PRAAT was accurately tracking the formants the outputs were compared with visual inspections of relevant spectrograms and power spectra (using cepstral smoothing: 400 Hz). Because PRAAT was unable to accurately track the lower three formants a second analysis was run. The only analysis parameters that changed were maximum formant value =2000 Hz and maximum number of formants=3. Formant frequency values for $F4$, $F5$, and $F6$ came from the first run, and $F1$, $F2$, and $F3$ came from the second run. The formant frequency values from both analyses were combined and mean formant spacing (ΔF) was estimated using a regression method in which each formant value is plotted against its expected value [this method is covered in more detail by Reby and McComb (2003)].

The linear regression method of Reby and McComb (2003) requires the expected formant positions to be plotted against actual measured values, and this requires a vocal tract model to provide the expected formant values to regress the observed values against. As giant panda bleats are delivered with a partially or fully closed mouth (Peters, 1985) but do not appear to be fully nasalized, we modeled the vocal tract as a tube closed at both ends (for more details, see the Appendix).

D. Analysis of relatedness

The pairwise genetic relatedness of individuals was quantified using their coefficient of relatedness (r). This coefficient estimates the degree to which two individuals share identical alleles, e.g., r=0.50 between full siblings and between parents and offspring, r=0.25 between half siblings and between grandparent and grandchild, etc. The 2003 In-

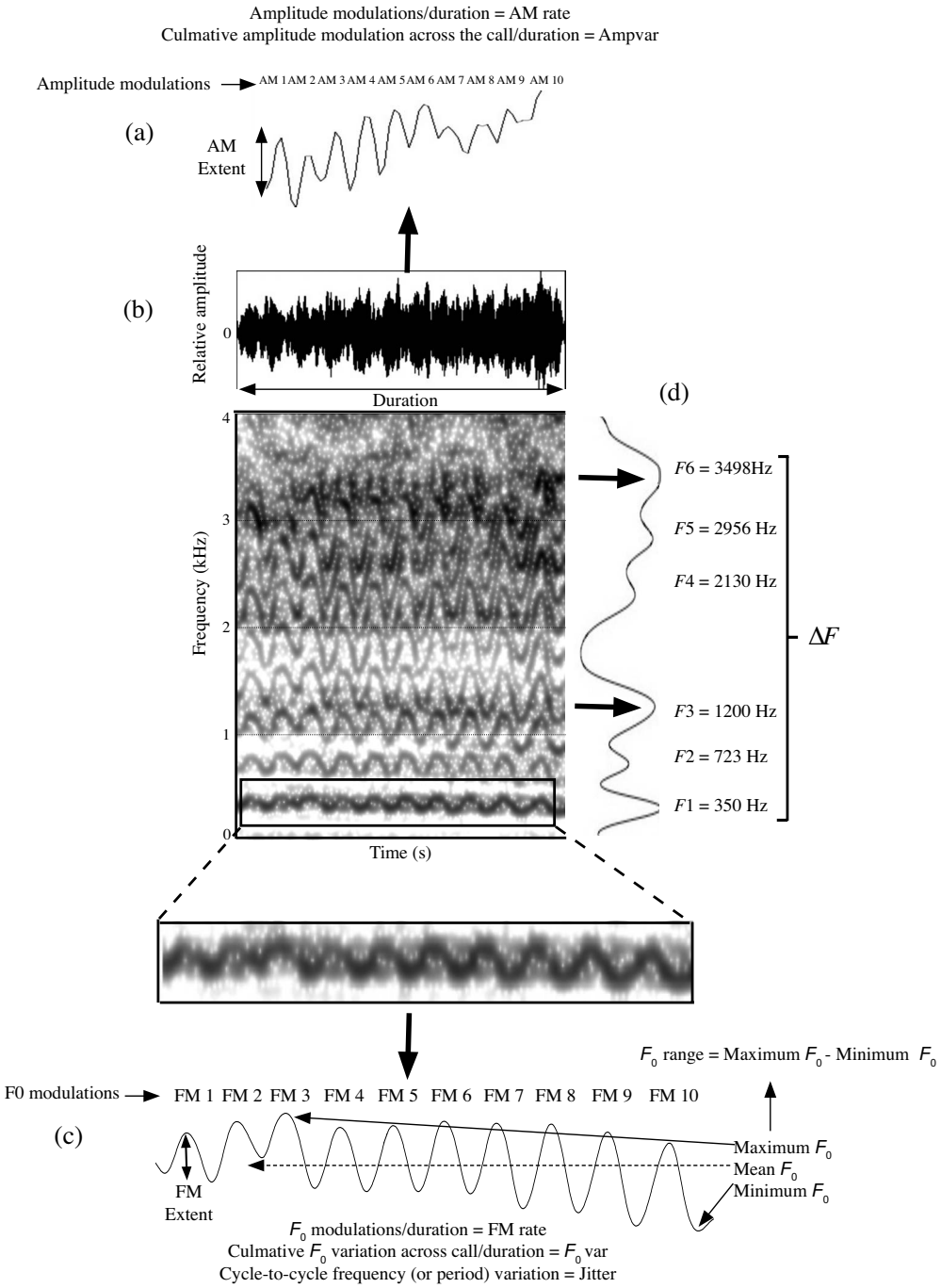


FIG. 2. Illustration of the source- and filter-related acoustic features made on (a) the intensity contour (Ampvar, AM extent, and AM rate); (b) the waveform (duration); (c) the F_0 contour (mean F_0, maximum F_0, minimum F_0, F_0 range, F_0 var, FM extent, FM rate, and jitter); and (d) LPC spectrum ($F1$, $F2$, $F3$, $F4$, $F5$, $F6$, and ΔF).

ternational Studbook for the Giant Panda was used to obtain information on the ancestry of all the subjects (Xie and Gipps, 2003). The r values could then be used to create a matrix of the pairwise genetic relatedness of our subjects (see Table II for r values of related subjects).

E. Statistical analyses

Although some of the acoustic measures in the analysis were correlated, the raw variables were retained to minimize loss of information. In order to assess individual differences in the acoustic structure of male bleats a discriminant function analysis (DFA) was used to classify giant panda bleats using subject identity as the group identifier and the acoustic variables as discriminant variables. For each classification, both the re-classification and the more conservative leave-one-out, cross-validation procedure was applied. Moreover, because collapsing data across sexes could falsely inflate classification results, classification percentages for both sexes are reported separately (following Bachorowski and Owren, 1999). In addition, the percentage error reduction associated with the classification accuracies is given. The percentage error reduction term takes into account the chance error rate and, hence, produces an unbiased measure of the

TABLE II. Confusion matrix to show percentage correct classification and misattributions for each subject. Where subjects are related their coefficient of relatedness (rounded up to two decimal places) is given in parentheses. Observed percentage of correct classification for each subject is given diagonally in bold. Correct percentage of re-classified bleats after leave-one-out, cross validation is also given for each subject. In addition, each subject's percentage representation in the sample of 245 bleats is given. A partial Mantel test (controlling for subject age) indicated that bleats were not consistently misattributed to more closely related individuals (r=0.102, p=0.112).

Subject	Sex	Representation in sample (%)	Correct re-classification (%)	Bing bing	Bing dian	Cheng gong	Cheng ji	Da shuang	Kebi	Le le	Li li	Lin lin	Lun lun	Ping ping	Qi yuan	Qi zhen	Shu lan	Tian tian	Xiong bang	Xiao shuang	Yang yang
Bing bing	Female	4	30	**70**	0(0.50)	0(0.13)	20(0.13)	0(0.13)	0	0	0(0.13)	0	0(0.50)	0	0	0	10	0	0	0(0.13)	0
Bing dian	Male	12	100	0(0.50)	**100**	0(0.07)	0(0.07)	0(0.25)	0	0	0(0.07)	0	0(0.25)	0	0	0	0	0	0	0(0.25)	0
Cheng gong	Female	4	60	20(0.13)	0(0.07)	**80**	0(0.50)	0(0.13)	0(0.50)	0	0(0.13)	0	0(0.07)	0	0(0.07)	0(0.07)	0(0.13)	0	0(0.07)	0(0.13)	0
Cheng ji	Female	5	78.6	0(0.13)	0(0.07)	7.1(0.50)	**92.9**	0	0(0.50)	0	0(0.13)	0	0(0.07)	0	0(0.07)	0(0.07)	0(0.13)	0	0(0.07)	0(0.13)	0
Da shuang	Female	4	61.5	0(0.13)	0(0.25)	0(0.13)	0	**84.6**	0	0	0(0.25)	0	7.7(0.07)	0	7.7(0.25)	0(0.25)	0	0	0	0(0.50)	0(0.13)
Kebi	Male	6	77.8	0	0	0(0.50)	0(0.50)	0	**83.3**	0	0	0	0	0	0(0.13)	0(0.13)	0(0.25)	16.7	0(0.13)	0	0
Le le	Male	3	71.4	0	0	0	0	0	0	**100**	0	0	0	0	0	0	0	0	0	0	0
Li li	Female	5	76.9	0(0.13)	0(0.07)	0(0.13)	0(0.13)	0(0.50)	15.4	0	**84.6**	0	0(0.07)	0	0	0	0	0	0	0(0.25)	0(0.25)
Lin lin	Male	7	77.8	0	0	0	0	0	0	0	0	**88.9**	0	0	0	0	0	11.1(0.25)	0	0	0
Lun lun	Female	3	71.4	0(0.50)	0(0.25)	0(0.07)	0(0.07)	0(0.50)	0	0	0(0.07)	0	**100**	0	0(0.13)	0(0.13)	0(0.25)	0	0	0(0.07)	0
Ping ping	Male	9	73.7	15.8	0	0	0	0	0	0	0	0	0	**73.7**	0	0	5.3	10.5	0	0	0
Qi yuan	Female	5	73.3	0	0	0(0.07)	0(0.07)	0(0.25)	0(0.13)	0	0	0	0(0.13)	6.7	**86.7**	6.7	0(0.25)	0	0(0.25)	0(0.25)	0
Qi zhen	Female	4	55.6	0	0	0(0.07)	0(0.07)	22.2(0.25)	0(0.13)	0	0	0	0(0.13)	0	11.1	**66.7**	0(0.25)	0	0(0.25)	0(0.25)	0
Shu lan	Female	6	25	0	0	0(0.13)	0(0.13)	0	0(0.25)	0	0	0	0(0.25)	25	0(0.25)	0(0.25)	**62.5**	0	12.5(0.25)	0	0
Tian tian	Male	3	47.6	4.8	0	0	4.8	0	0	0	0	9.5(0.25)	0	4.8	0	0	4.8	**71.4**	4.8	0	0
Xiong bang	Male	8	50	0	0	0(0.07)	0(0.07)	0	0(0.13)	0	0	0	0	12.5	0(0.25)	0(0.25)	0(0.25)	0	**87.5**	0	0
Xiao shuang	Male	6	79.3	0(0.13)	0(0.25)	0(0.13)	0(0.07)	3.4(0.50)	0	0	0(0.25)	0	3.4(0.07)	3.4	0(0.25)	0(0.25)	0	0	0	**86.2**	3.4(0.13)
Yang yang	Male	6	86.7	0	0	0	0	0(0.13)	0	0	0(0.25)	0	0	0	6.7	0	0	6.7	0	0(0.13)	**86.7**

TABLE III. Descriptive statistics for the 19 acoustic measures used.

	Overall (N=18)				Males (N=9)				Females (N=9)			
Acoustic measures	M	s.d.	Minimum	Maximum	M	s.d.	Minimum	Maximum	M	s.d.	Minimum	Maximum
Duration	1.08	0.20	0.79	1.64	1.00	0.14	0.79	1.29	1.15	0.23	0.97	1.64
Mean F_0	365.4	51.7	298.6	513.8	371.9	62.7	315.7	513.8	358.9	40.6	298.6	419.1
Minimum F_0	280.9	19.4	258.2	327.1	278.8	17.4	260.9	304.4	283.0	22.1	258.2	327.1
Maximum F_0	493.1	59.0	379.1	625.5	509.0	60.2	427.5	625.5	477.2	56.5	379.1	528.4
F_0 range	212.2	48.8	120.9	323.2	230.1	46.9	165.3	323.2	194.2	46.2	120.9	248.2
F_0 var	1898.3	560.2	749.6	3167.4	2108.4	570.6	1106.7	3167.4	1688.3	492.0	749.6	2393.8
FM rate	9.51	1.24	7.36	11.4	9.16	1.30	7.36	11.4	9.87	1.14	7.59	11.4
FM extent	100.7	27.5	48.9	157.5	116.2	25.3	76.0	157.5	85.3	20.9	48.9	122.7
Jitter	4.35	0.76	2.94	5.89	4.79	0.61	4.04	5.89	3.91	0.65	2.94	5.18
Ampvar	110.7	41.3	75.6	258.7	126.3	52.5	87.8	258.7	95.0	17.8	75.6	132.9
AM rate	8.64	0.89	7.14	11.2	8.85	1.16	7.14	11.2	8.43	0.51	7.54	8.95
AM extent	13.1	5.13	9.16	31.1	14.7	6.62	10.0	31.1	11.56	2.53	9.16	17.77
$F1$	397.4	51.5	324.9	556.4	385.5	35.4	326.2	446.6	409.3	63.7	324.9	556.4
$F2$	733.8	58.7	629.4	835.4	728.3	69.2	629.4	835.4	739.3	49.6	657.7	806.7
$F3$	1269.1	60.7	1190.1	1438.7	1251.1	48.4	1190.1	1350.7	1287.1	68.9	1230.4	1438.7
$F4$	2191.2	67.2	2073.1	2294.9	2185.8	74.7	2073.1	2294.9	2196.5	63.0	2107.0	2279.1
$F5$	3048.0	100.0	2912.4	3246.4	3005.0	69.2	2912.4	3094.3	3091.1	110.9	2914.3	3246.4
$F6$	3565.1	75.9	3420.8	3729.4	3529.1	66.0	3420.8	3633.4	3601.0	70.4	3481.8	3729.4
ΔF	561.1	9.48	543.3	577.0	555.3	8.23	543.3	567.6	566.9	6.89	558.3	577.0

level of correct classification (Bachorowski and Owren, 1999). Because of uneven subject participation in the data set the percentage correct classification expected due to chance was calculated according to the group sizes. The statistical significance of correct classification of bleats to each individual and across all subjects was obtained using the chi square statistic.

To assess the influence of genetic relatedness on the acoustic structure of bleats acoustic measures were transformed into Z scores, and dissimilarity matrixes of Euclidean distances between individuals were generated, first as a measure of overall call similarity (entering all the acoustic measures together) and then for each acoustic measure separately. Correspondence between pairwise individual relatedness (created using the coefficient of relatedness) and these dissimilarity matrices was then tested using partial Mantel tests (Mantel, 1967). A partial Mantel test was also used to see if bleats were consistently misattributed to more closely related individuals by the DFA (see Table II). Mantel tests are permutation procedures that are ideal for testing the statistical significance of correlations between matrices of the same rank that exhibit linear correlation between variables and independent observation pairs (Sokal and Rohlf, 1995). Because age may affect the acoustic structure of a caller's vocalizations (Reby and McComb, 2003; Sanvito *et al.*, 2007) partial Mantel tests controlling for subject age were used. In addition, the step-up-Hochburg technique was used to correct for multiple testing (Pfefferle and Fischer, 2006). This algorithm sequentially adjusts the p values required to attain statistical significance; the first critical value is as in a Bonferrroni correction (alpha=$0.05/n$) where n =the number of tests, the second=$2\times(0.05/n)$, the third =$3\times(0.05/n)$, etc. Mantel tests were computed using XLSTAT 2008 (Addinsoft, NY). Other analyses were conducted using SPSS version 15 and significance levels were set at 0.05.

III. RESULTS

A. Acoustic structure of giant panda bleats

Descriptive statistics for all source- and filter-related features of giant panda bleats both across and within each sex are given in Table III. Giant panda bleats ranged up to approximately 1.6 s in duration and had mean F_0 values that varied from approximately 300 Hz to more than 500 Hz. The minimum F_0 was much less variable, only differing by approximately 100 Hz between minimum and maximum values, whereas the maximum F_0 ranged from approximately 380 to over 600 Hz. The mean F_0 range was around 210 Hz and varied from approximately 120 to 320 Hz. The variability of F_0 along the call was quantified using the cumulative variation in the F_0 contour per second (F_0 var), the number of complete cycles of F_0 modulation per second (FM rate), and the mean peak-to-peak variation of each F_0 modulation (FM extent). F_0 var was highly variable and ranged from around 750 to 3170 Hz, the FM rate was approximately 9.5 cps, and mean FM extent was approximately 100 Hz. Short-term variability in F_0 was quantified using jitter and this varied by less than 3% across subjects. Amplitude variation was quantified using the cumulative variation in amplitude per second (Ampvar), the number of complete cycles of amplitude modulation per second (AM rate), and the mean peak-to-peak variation of each amplitude modulation (AM extent). Ampvar was around 111 dB and quite variable, ranging from approximately 75 to 260 dB; AM rate was approximately 8.5 cps; and AM extent was around 13 dB and varied from approximately 9 to 31 dB. Finally, there were six ob-

TABLE IV. Tests of equality of group means among individuals for each of the acoustic features used in the DFA. In all cases $p<0.001$.

Acoustic measures	Wilks' lambda	F
Duration	0.79	3.39
Mean F_0	0.22	46.8
Maximum F_0	0.53	11.7
Minimum F_0	0.46	15.9
F_0 range	0.61	8.43
F_0 var	0.61	8.73
FM rate	0.61	8.51
FM extent	0.62	8.03
Jitter	0.65	7.31
Ampvar	0.26	38.4
AM rate	0.69	5.98
AM extent	0.29	33.5
$F1$	0.56	10.7
$F2$	0.64	7.49
$F3$	0.75	4.41
$F4$	0.54	11.3
$F5$	0.39	21.1
$F6$	0.55	11.1
ΔF	0.45	16.3

servable and stable formants in the frequency range 0–3800 Hz (see Fig. 2) and the mean formant spacing (ΔF) was around 560 Hz.

B. Individual differences in the acoustic structure of bleats

The DFA correctly classified 83.7% of 245 bleats to the 18 individuals (an error reduction of 82.7%), falling to 69.0% (an error reduction of 67.6%) when a more conservative leave-one-out, cross validation was applied (see Table II for percentage correct classification for each individual). Because correct classification of bleats to individuals was higher when both sexes were analyzed separately (males: 89.9% and 79.9% cross-validated; females: 91.5% and 73.6% cross-validated), it appears that collapsing data across sexes did not falsely inflate these classification results. Compared to that expected by chance the level of classification across sexes was statistically significant for each individual and across all individuals ($p<0.001$). In addition, the univariate analysis showed that all the acoustic features measured differed significantly between individuals (see Table IV: all $p<0.001$). The structure matrix generated by the multivariate DFA (Table V) shows that the main contributors to individual vocal distinctiveness were mean F_0, Ampvar, and AM extent, followed by $F5$ and ΔF (see Table V for more information on the variance explained by each of the first four discriminant factors with eigenvalues >1 and the loading of each acoustic measure on these factors).

C. Acoustic cues to individual relatedness in bleats

Pairwise individual relatedness (obtained using coefficients of relatedness) was not correlated with overall acoustic similarity (see Table VI). In addition, the percentage of bleats incorrectly classified to each individual by the DFA was not correlated with overall acoustic similarity ($r=0.102$, $p=0.112$), indicating that misattributions were not more likely to occur among closely related individuals (see Table II). When each acoustic measure was considered separately, pairwise individual relatedness was correlated with AM rate and F_0 range (see Table VI). Both these acoustic characteristics were more similar (i.e., had lower dissimilarity) between more closely related individuals. After step-up-Hochburg adjustments, no other acoustic characteristics were significantly correlated with individual relatedness (see Table VI).

TABLE V. DFA structure matrix showing pooled within-groups correlations among discriminating variables and the first four standardized canonical discriminant functions with eigenvalues >1. Only correlations between each variable and any discriminant function >0.4 are shown.

	Discriminant functions			
Acoustic measures	1	2	3	4
Mean F_0	0.66			
Ampvar	0.44	0.56		
AM extent		0.57		
$F5$		−0.48		
ΔF			−0.52	
$F6$				0.54
$F1$			−0.41	
Eigenvalue	7.07	4.07	2.17	1.38
% of variance	40.0	23.0	12.2	7.8
Cumulative%	40.0	63.0	75.2	83.0

IV. DISCUSSION

A. Acoustic structure of giant panda bleats

This study provides a quantitative description of the acoustic structure of male and female giant panda bleats. The

TABLE VI. Pearson's correlation coefficients and p values for partial Mantel results (controlling for age) to illustrate the influence of genetic relatedness on the acoustic structure of bleats. Significant correlations before step-up-Hochburg adjustments are in bold, and * denotes significant after step-up-Hochburg adjustments for multiple statistical testing.

Acoustic measures	r	p
Duration	0.019	0.721
Mean F_0	0.036	0.519
Maximum F_0	0.087	0.128
Minimum F_0	−0.033	0.567
F_0 range	−0.176	**0.003***
F_0 var	0.112	0.058
FM rate	−0.089	0.110
FM extent	−0.148	**0.013**
Jitter	−0.020	0.735
Ampvar	0.041	0.455
AM rate	−0.226	**<0.001***
AM extent	0.069	0.247
$F1$	0.061	0.294
$F2$	−0.043	0.376
$F3$	−0.020	0.746
$F4$	−0.115	**0.041**
$F5$	0.048	0.419
$F6$	−0.143	**0.018**
ΔF	0.053	0.347
Overall (all measures)	0.062	0.285

acoustic data described here report 19 acoustic measures of giant panda bleats both across and within sexes (see Table III) and, hence, is more extensive than that of previous studies on this species (Peters, 1982, 1985; Charlton *et al.*, in press). Previous work revealed that male giant panda bleats had lower ΔF, higher jitter, and higher FM extent than female bleats (Charlton *et al.*, in press), and the descriptive statistics of the current study accord well with these known sex differences. It is interesting to note that despite the sexual size dimorphism in this species (Schaller *et al.*, 1985), no significant differences in mean, minimum, and maximum F_0 values exist between the sexes (Charlton *et al.*, in press). Finally, giant panda bleats are also characterized by rapid amplitude and F_0 modulation and have stable, flat formants (see Fig. 2). This allows the formant frequencies to be distinguished from the rapidly modulated F_0 and its related harmonics, and for these filter-related acoustic features to be correctly identified and measured.

B. Individual differences in the acoustic structure of bleats

The DFA confirmed that giant panda bleats are individually distinctive, with a high level of correct classification (69%) achieved when the more conservative leave-one-out, cross-validation approach was used. Discriminant analyses were also conducted within sexes, because the known sex differences in acoustic parameters of giant panda bleats (Charlton *et al.*, in press) could have produced distinct patterns of sorting accuracy across sexes. Interestingly, bleats were correctly classified to individuals at higher levels when both sexes were considered separately (males: 79.9% cross-validated; females: 73.6% cross-validated). While this indicates that collapsing data across sexes did not falsely inflate the initial classification results it represents an unusual finding. The most conservative conclusion for the higher within sex classification accuracies reported here is that relatively few acoustic differences in giant panda bleats exist between the sexes (Charlton *et al.*, in press), in comparison to human speech, for example (Bachorowski and Owren, 1999), and that it is easier for the model to correctly classify bleats to individuals when there are fewer subjects and bleats in the analysis (139 male bleats and 106 female bleats to classify to 9 individuals in each case instead of 245 bleats to $N=18$).

Across sexes, the acoustic features contributing most to individual distinctiveness were mean F_0, Ampvar, and AM extent, followed by ΔF and FM extent. Previous work on acoustic individuality in mammals, in which the biomechanical modes of production were considered, found source-related features (i.e., those produced by the larynx) to be the most highly distinctive (McComb *et al.*, 2003; Vannoni and McElligott, 2006) whereas others found filter-related features (i.e., those produced by the filtering affect of the vocal tract) to be more individually distinctive (Rendall, 2003; Soltis *et al.*, 2005). The results presented here suggest that source-related features, in particular, mean F_0 and features relating to amplitude modulation, contribute the most to vocal identity in giant pandas.

Mean F_0 is highly individualized in several mammal species (Charrier *et al.*, 2003; Searby and Jouventin, 2003; Torriani *et al.*, 2006). This is not surprising because the range of possible fundamental frequencies a mammal can produce is ultimately constrained by factors such as the length and stiffness of the vocal folds (Titze, 1994), and these factors are likely to vary among individuals. In this study, mean F_0 was highly individually distinctive whereas the minimum and maximum F_0 were somewhat less so. One possible explanation is that F_0 modulation results in more stabilized vocal fold vibration patterns, providing a narrower range of mean F_0 within an individual and, hence, increasing the potential for this acoustic feature to encode identity. Indeed, a narrower F_0 range would also prevent minimum and maximum F_0 from reflecting the limits of vocal fold vibration, and in doing so reduce their potential for coding individuality. In support of this explanation, the F_0 modulated bleat vocalizations of sheep (*Ovies ovies*) also have highly individualized mean F_0 that are more individually distinctive than the minimum and maximum F_0 of their vocalizations (Searby and Jouventin, 2003).

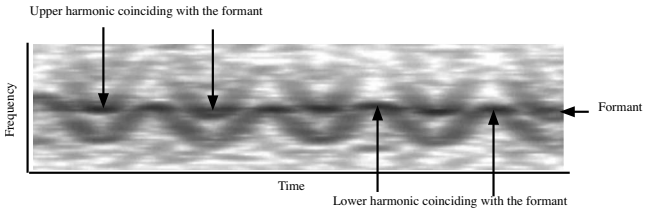


FIG. 3. Section of a giant panda bleat spectrogram showing the third and fourth harmonics (labeled upper and lower) of the fundamental frequency striking *F*3. As a harmonic coincides with the center frequency of the formant the darkening on the spectrogram indicates an increase in frequency amplitude.

Although amplitude modulation contributes to vocal distinctiveness and recognition in animals (Aubin and Jouventin, 2002) it is less clear why measures of amplitude variation should be so individually distinctive. Amplitude can be regulated by varying sub-glottal air pressure, altering the glottal width affecting the amount of aerodynamic power that is converted to acoustic power and/or when harmonics of the F_0 coincide with formants (Titze, 1994). During F_0 modulation the harmonics sweep back and forth across the formants, causing their individual amplitude levels to fluctuate accordingly. This rhythmic attenuating and enhancing of harmonic amplitude levels as they coincide with formants is thought to be the main cause of amplitude modulation in human vibrato (Horii and Hata, 1988). Observations of giant panda bleat spectrograms show harmonics of the strongly modulated F_0 consistently striking formants and yielding amplitude peaks, suggesting that this type of phenomenon also occurs in this species (see Fig. 3). Moreover, the strong correlations between amplitude modulation per second (Ampvar) and both F_0 var ($r=0.164$, $p=0.01$) and FM rate ($r=0.201$, $p<0.01$) lend further support to this contention.

Furthermore, if we consider that the amount of amplitude modulation produced in this way is affected by the harmonic spacing (the F_0 at any given point), the rate and extent of F_0 modulations, and the spacing of the formants, the amount of amplitude variation and mean extent of amplitude modulation across a bleat may become highly individualized

simply because it brings together information on several acoustic features, which are themselves individually distinctive. Indeed, after mean F_0, Ampvar, and AM extent, FM extent and ΔF were the next most individually distinctive. Given that F_0 modulation improves vowel perception in humans by increasing the salience of the formants (Ryalls and Lieberman, 1982), the characteristic frequency modulation of giant panda bleats may even have evolved to facilitate vocal recognition by highlighting a caller's distinctive formant pattern.

Indeed, other studies on humans and other nonhuman mammals have emphasized the importance of vocal tract resonances as cues to individual identity (Owren *et al.*, 1997; Bachorowski and Owren, 1999; Reby *et al.*, 2006). In particular, formants are likely to be a reliable source of identity cues in large bodied mammals that do not perform dynamic vocal tract modifications during call production (so that the formants are comparatively stable within individuals across calls) and in which the source energy of the call adequately highlights the caller's distinctive formant pattern (Rendall *et al.*, 1998; Owren and Rendall, 2003; Rendall, 2003). Because the formants in giant panda bleats are stable (see Figs. 1 and 2), and therefore consistent among the calls an individual makes, they are likely to be highly individually distinctive. Moreover, the dense harmonic structure and F_0 modulation should highlight the formants, making giant panda bleats ideal for identity cueing [as discussed by Owren and Rendall (1997, 2001)].

Although the formants of giant panda bleats may have been expected to feature more highly when classifying individuals to their vocalizations., they did still contribute strongly to individual vocal distinctiveness, with ΔF, and $F5$ being the most individually distinctive (see Table V). Indeed, the high overall level of bleat classification to individuals observed may have masked the individual distinctiveness of filter-related features, which are still individually distinctive (see Table IV) but just relatively less so than source-related features. Because the formants were correctly identified and measured this could not have been a factor contributing to the relatively low individuality of filter-related features when compared to source-related features. However, formant spacing may be a relatively poor cue to identity in giant pandas because it depends on the caller's vocal tract length (VTL), which could overlap among individuals. This is particularly likely in giant pandas because they have low variability in body size (Janssen *et al.*, 2006) and a relatively stereotyped body position during vocalization (personal observation). Moreover, while formants passively contribute to individual vocal identity because they are explicitly linked to the caller's VTL and shape, there may have been additional selection for source-related acoustic features to become more individually distinctive than they would otherwise have been through natural variation in laryngeal morphology.

C. Acoustic cues to individual relatedness in bleats

Currently we know very little about how genetics mediates the development and expression of acoustic structure in mammals. Although closely related individuals might be expected to have similarities in call structure simply due to physical similarities in vocal production anatomy, differing environmental and hormonal factors could confound any genetically mediated similarities in vocal characteristics among individuals that may otherwise exist. For example, estrogen and testosterone are both known to affect source-related vocal characteristics (Abitbol *et al.*, 1999; Dabbs and Mallinger, 1999; Fuchs *et al.*, 2007), and variation in the levels of these sex hormones, either seasonal or during development, could obscure inherent similarities among closely related individuals both across and within sexes. Indeed, no evidence of acoustic similarity among closely related individuals was found when all of the acoustic measures were considered in the analysis (see Table VI). In addition, it does not appear that bleats were consistently misattributed to more closely related individuals (see Table II). However, after correcting for multiple statistical testing, AM rate and F_0 range were predictive of genetic relatedness and, hence, these acoustic features may be heritable components of giant panda bleats that could be used as a measure of genetic relatedness in this species.

If we consider that amplitude modulation in giant panda bleats is in part generated when harmonics of the F_0 coincide with formants, then a correlation between this acoustic feature and relatedness may have been predicted because it brings together information on source and filter acoustic features that are likely to reflect both a given callers' laryngeal and vocal tract morphology (see earlier discussion). In addition, F_0 range is likely to be at least partly constrained by vocal fold length (Titze, 1994), which may also be a heritable feature. However, it must be noted that our measure of genetic relatedness estimates the amount of genes shared by individuals, and is therefore only an approximate guide to genetic similarity. Future studies should use DNA fingerprinting techniques (which were unavailable to us) to accurately determine the number of alleles shared between individuals, in order to explore the precise relationship between genetic similarity and vocal characteristics in this species.

V. CONCLUSIONS

The results presented here show that giant panda bleats are individually distinctive and indicate that source-related features, in particular, mean F_0, Ampvar, and AM extent, contribute the most to vocal identity. In addition, although individual pairwise relatedness was not correlated with overall acoustic similarity, it was highly correlated with AM rate and F_0 range, suggesting that these acoustic features are heritable components of giant panda bleats that could be used as a measure of genetic relatedness.

The high level of inter-individual variability in the acoustic structure of giant panda bleats suggests that they may function as vocal signatures in this species vocal communication system. Moreover, interactions with other conspecifics increase considerably during the breeding season when vocal distinctiveness may have fitness benefits for male and female giant pandas in the contexts of intra-sexual competition (Barnard and Burk, 1979; Tripovich *et al.*, 2008) and mate choice (East and Hofer, 1991). The production of

individually distinctive scent marks (Hagey and Macdonald, 2003) and the ability to discriminate between individuals on the basis of these olfactory signals (Swaisgood *et al.*, 1999) suggests that individual recognition is important for giant pandas. Whether giant pandas can recognize specific individuals by their bleats or even distinguish among the bleats of different individuals would have to be proved unequivocally; however, the high degree of individual distinctiveness we found in the acoustic structure of bleats would make this a strong possibility.

The functional relevance of different acoustic characteristics in this species' sexual communication and whether giant pandas can discriminate between different callers requires testing with playback experiments. Nevertheless, given that giant panda bleats are individually distinctive, these findings suggest that the acoustic structure of these vocalizations might be used as a bioacoustic tool to assess population levels in this species. This type of information is critical for wildlife conservation, and establishing practical monitoring protocols that are non-invasive and cost effective is particularly challenging for this elusive mammal that inhabits densely forested and remote mountain areas. We suggest that animal activated cameras that also capture sound could be set up at known giant panda breeding sites and communal scent stations. Given the high rates of giant panda vocal activity during the breeding season it may be possible to obtain enough good quality recordings for bioacoustic techniques to be used to help distinguish between different individuals, allowing the size of a given population to be estimated, and even to track individuals and thus dispersal in this highly endangered species.

ACKNOWLEDGMENTS

We thank China's Ministry of Construction for its support and cooperation and all the staff at Chengdu Research Base of Giant Panda Breeding in Chengdu, Sichuan Province, People's Republic of China. We would also like to thank the staff from Zoo Atlanta, National Zoological Park, and Memphis Zoo for allowing B.D.C. to record their giant pandas. Finally, we would also like to thank Michael Owren, David Reby, and an anonymous reviewer for their helpful and insightful comments on an earlier version of the manuscript, and my brother Andrew Charlton for helping to run some of the statistical tests. This material is based on work supported in part by the Center for Behavioral Neuroscience under the STC Program of the National Science Foundation under Agreement No. IBN-9876754.

APPENDIX

Initially, because no data on giant panda vocal tract length (VTL) were available, the formants in giant panda bleats were visualized on spectrograms and power spectra, and not measured with *a priori* expectations on the number of resonances. The resonances identified in this way were always static across bleats (see Fig. 2) making it clear that they could not be harmonics of the strongly modulated F_0. Only after obtaining a radiograph of a male subject's head and neck region in 2008 were we able to measure the distance from teeth to glottis using a flexible ruler, and obtain a known VTL of around 32 cm. This known measurement could then be compared with the subject's VTL estimated from recordings made during the same year, using the equation $\mathrm{VTL}=c/2\Delta F$ where c is the approximate speed of sound in the mammalian vocal tract (350 m/s) (Titze, 1994) and ΔF is the formant spacing calculated using the regression method of Reby and McComb (2003).

The commonly used "one-side-open" tube model to determine formant spacing gave an estimated VTL of around 25 cm, 6 cm shorter than this male subject's actual VTL of 32 cm. In contrast, the "closed-both-ends" tube model returned formant spacing values that gave an estimated VTL of 31.79 cm for this individual. Although the uniform tube model is only a rough approximation for estimating the number of formants expected, the closed mouth posture adopted by giant pandas when bleating and the close correspondence between a male subject's actual and estimated VTL make the closed-both-ends tube model the most appropriate way to model the giant panda vocal tract. It must also be noted that the closed-both-ends tube model yields the same predicted formant values and number of formants as a tube open at both ends (Titze, 1994). If we consider that acoustic energy is exiting via the nostrils in giant pandas, which constitute a significantly smaller exit hole than the mouth, and that the glottis is, in fact, partially open, the closed-both-ends model seems even more appropriate to use.

After a workable model had been established predictions could be made about the number of formants to expect in a given frequency range. For a linear tube closed at both ends, six resonances would be expected in the frequency range 0–3800 Hz ($F1=c/2\times\mathrm{VTL}$, $F2=2\times F1$, $F3=3\times F1$, etc.). Accordingly, in this study, the automated programs in PRAAT could then be set to track and measure six formants in the frequency range 0–3800 Hz. Although the giant panda vocal tract clearly departs from the linear tube model (for example, only one formant at ~2200 Hz appears in the frequency range 1300–3000 Hz: see Fig. 2) the closed-both-ends model yielded overall formant spacing that gave estimated VTLs corresponding to our anatomical data on VTL. Moreover, if one of the formants had been incorrectly identified then there would be fewer identifiable formants in the bleats, and this, by increasing the formant spacing, would then yield an unrealistically short apparent VTL.

Abitbol, J., Abitbol, P., and Abitbol, B. (**1999**). "Sex hormones and the female voice," J. Voice **13**, 424–446.

Aubin, T., and Jouventin, P. (**2002**). "How to vocally identify kin in a crowd: The penguin model," Adv. Study Behav. **31**, 243–277.

Bachorowski, J. A., and Owren, M. J. (**1999**). "Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech," J. Acoust. Soc. Am. **106**, 1054–1063.

Barnard, C. J., and Burk, T. (**1979**). "Dominance hierarchies and the evolution of individual recognition," J. Theor. Biol. **81**, 65–73.

Blumstein, D. T., and Munos, O. (**2005**). "Individual, age and sex-specific information is contained in yellow-bellied marmot alarm calls," Anim. Behav. **69**, 353–361.

Boersma, P., and Weenink, D. (**2005**). PRAAT: Doing phonetics by computer (version 4.3.01) [computer program], retrieved from http://www.praat.org/ (Last viewed 7/20/09).

Charlton, B. D., Zhihe, Z., and Snyder, R. J. (**2009**). "The information content of giant panda (*Ailuropoda melanoleuca*) bleats: Acoustic cues to sex, age and size," Anim. Behav. **78**, 893–898.

Charrier, I., Mathevon, N., and Jouventin, P. (**2001**). "Mother's voice recognition by seal pups—Newborns need to learn their mother's call before she can take off on a fishing trip," Nature (London) **412**, 873.
Charrier, I., Mathevon, N., and Jouventin, P. (**2003**). "Individuality in the voice of fur seal females: An analysis study of the pup attraction call in *Arctocephalus tropicalis*," Marine Mammal Sci. **19**, 161–172.
Crockford, C., Herbinger, I., Vigilant, L., and Boesch, C. (**2004**). "Wild chimpanzees produce group-specific calls: A case for vocal learning?," Ethology **110**, 221–243.
Dabbs, J. M., and Mallinger, A. (**1999**). "High testosterone levels predict low voice pitch among men," Pers. Individ. Differ. **27**, 801–804.
East, M. L., and Hofer, H. (**1991**). "Loud calling in a female-dominated mammalian society. 2. Behavioral contexts and functions of whooping of spotted hyenas, *Crocuta crocuta*," Anim. Behav. **42**, 651–669.
Egnor, S. E. R., and Hauser, M. D. (**2004**). "A paradox in the evolution of primate vocal learning," Trends Neurosci. **27**, 649–654.
Fant, G. (**1960**). *Acoustic Theory of Speech Production* (MOUTON, The Hague).
Fitch, W. T. (**1997**). "Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques," J. Acoust. Soc. Am. **102**, 1213–1222.
Fitch, W. T., and Hauser, M. D. (**1995**). "Vocal production in nonhuman-primates—Acoustics, physiology, and functional constraints on honest advertisement," Am. J. Primatol **37**, 191–219.
Fitch, W. T., and Hauser, M. D. (**2002**). "Unpacking 'honesty': Generating and extracting information from acoustic signals," in *Animal Communication*, edited by A. Megala-Simmons and A. Popper (Springer-Verlag, Berlin).
Frommolt, K. H., Goltsman, M. E., and MacDonald, D. W. (**2003**). "Barking foxes, *Alopex lagopus*: Field experiments in individual recognition in a territorial mammal," Anim. Behav. **65**, 509–518.
Fuchs, M., Froehlich, M., Hentschel, B., Stuermer, I. W., Kruse, E., and Knauft, D. (**2007**). "Predicting mutational change in the speaking voice of boys," J. Voice **21**, 169–178.
Hagey, L., and Macdonald, E. (**2003**). "Chemical cues identify gender and individuality in giant pandas (*Ailuropoda melanoleuca*)," Chem. Ecol. **29**, 1479–1488.
Hare, J. F. (**1998**). "Juvenile Richardson's ground squirrels, *Spermophilus richardsonii*, discriminate among individual alarm callers," Anim. Behav. **55**, 451–460.
Harris, T. R., Fitch, W. T., Goldstein, L. M., and Fashing, P. J. (**2006**). "Black and white colobus monkey (*Colobus guereza*) roars as a source of both honest and exaggerated information about body mass," Ethology **112**, 911–920.
Horii, Y., and Hata, K. (**1988**). "A note on phase-relationships between frequency and amplitude modulations in vocal vibrato," Folia Phoniatr (Basel) **40**, 303–311.
Insley, S. J. (**2000**). "Long-term vocal recognition in the northern fur seal," Nature (London) **406**, 404–405.
Janik, V. M., and Slater, P. J. B. (**1997**). "Vocal learning in mammals," Adv. Study Behav. **26**, 59–99.
Janssen, D. L., Edwards, M., Sutherland-Smith, M., Yu, J., Li, D., Zhang, G., Wei, R., Zhang, C. L., Miller, R. E., Phillips, L. G., Hu, D., and Tang, C. (**2006**). "Significant medical issues and biological reference values for giant pandas from the biomedical survey," in *Giant Pandas: Biology, Veterinary Medicine and Management*, edited by D. E. Wildt, A. Zhang, H. Zhang, D. J. Janssen, and S. Ellis (Cambridge University Press, New York), pp. 59–86.
Kleiman, D. G. (**1983**). "Ethology and reproduction of captive giant pandas (*Ailuropoda melanoleuca*)," Z. Tierpsychol **62**, 1–46.
Kleiman, D. G., Karesh, W. B., and Chu, P. R. (**1979**). "Behavioural changes associated with oestrus in the giant panda *Ailuropoda melanoleuca*)," Int. Zoo Yearb. **19**, 217–223.
Lindburg, D. G., Czekala, N. M., and Swaisgood, R. R. (**2001**). "Hormonal and behavioral relationships during estrus in the giant panda," Zoo Biol. **20**, 537–543.
Mantel, N. (**1967**). "The detection of disease clustering and a generalized regression approach," Cancer Res. **27**, 209–220.
McComb, K., Moss, C., Sayialel, S., and Baker, L. (**2000**). "Unusually extensive networks of vocal recognition in African elephants," Anim. Behav. **59**, 1103–1109.
McComb, K., Reby, D., Baker, L., Moss, C., and Sayialel, S. (**2003**). "Long-distance communication of acoustic cues to social identity in African elephants," Anim. Behav. **65**, 317–329.
McCowan, B., and Hooper, S. L. (**2002**). "Individual acoustic variation in Belding's ground squirrel alarm chirps in the high Sierra Nevada," J. Acoust. Soc. Am. **111**, 1157–1160
McElligott, A. G., O'Neill, K. P., and Hayden, T. J. (**1999**). "Cumulative long-term investment in vocalization and mating success of fallow bucks, *Dama dama*," Anim. Behav. **57**, 1159–1167.
Owren, M. J., and Rendall, D. (**1997**). "An affect-conditioning model of nonhuman primate vocal signaling," in *Perspectives in Etology*, edited by D. H. Owings, M. D. Beecher, and N. S. Thompson (Plenum, New York), pp. 299–346.
Owren, M. J., and Rendall, D. (**2001**). "Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling," Evol. Anthropol. **10**, 58–71.
Owren, M. J., and Rendall, D. (**2003**). "Salience of caller identity in rhesus monkey (*Macaca mulatta*) coos and screams: Perceptual experiments with human (*Homo sapiens*) listeners," J. Comp. Psychol. **117**, 380–390.
Owren, M. J., Seyfarth, R. M., and Cheney, D. L. (**1997**). "The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cyncephalus ursinus*): Implications for production processes and functions," J. Acoust. Soc. Am. **101**, 2951–2963.
Peters, G. (**1982**). "A note on the vocal behaviour of the giant panda, *Ailuropoda melanoleuca* (David 1869)," Z. Saugetierkunde **47**, 236–246.
Peters, G. (**1985**). "A comparative survey of vocalization in the giant panda (*Ailuropoda melanoleuca*, David 1869)," Bongo (Berlin) **10**, 197–208.
Pfefferle, D., and Fischer, J. (**2006**). "Sounds and size: Identification of acoustic variables that reflect body size in hamadryas baboons, *Papio hamadryas*," Anim. Behav. **72**, 43–51.
Reby, D., Andre-Obrecht, R., Galinier, A., Farinas, J., and Cargnelutti, B. (**2006**). "Cepstral coefficients and hidden Markov models reveal idiosyncratic voice characteristics in red deer (*Cervus elaphus*) stags," J. Acoust. Soc. Am. **120**, 4080–4089.
Reby, D., Cargnelutti, B., Joachim, J., and Aulagnier, S. (**1999**). "Spectral acoustic structure of barking in roe deer (*Capreolus capreolus*). Sex-, age- and individual-related variations," C. R. Acad. Sci. III **322**, 271–279.
Reby, D., Hewison, M., Izquierdo, M., and Pepin, D. (**2001**). "Red deer (*Cervus elaphus*) hinds discriminate between the roars of their current harem-holder stag and those of neighbouring stags," Ethology **107**, 951–959.
Reby, D., Joachim, J., Lauga, J., Lek, S., and Aulagnier, S. (**1998**). "Individuality in the groans of fallow deer (*Dama dama*) bucks," J. Zool. **245**, 79–84.
Reby, D., and McComb, K. (**2003**). "Anatomical constraints generate honesty: Acoustic cues to age and weight in the roars of red deer stags," Anim. Behav. **65**, 519–530.
Rendall, D. (**2003**). "Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons," J. Acoust. Soc. Am. **113**, 3390–3402.
Rendall, D., Kollias, S., and Ney, C. (**2005**). "Pitch (Fo) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry," J. Acoust. Soc. Am. **117**, 944–955.
Rendall, D., Owren, M. J., and Rodman, P. S. (**1998**). "The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations," J. Acoust. Soc. Am. **103**, 602–614.
Rendall, D., Rodman, P. S., and Emond, R. E. (**1996**). "Vocal recognition of individuals and kin in free-ranging rhesus monkeys," Anim. Behav. **51**, 1007–1015.
Riede, T., Bronson, E., Hatzikirou, H., and Zuberbuhler, K. (**2005**). "Vocal production mechanisms in a non-human primate: Morphological data and a model," J. Hum. Evol. **48**, 85–96.
Ryalls, J. H., and Lieberman, P. (**1982**). "Fundamental-frequency and vowel perception," J. Acoust. Soc. Am. **72**, 1631–1634.
Sanvito, S., Galimberti, F., and Miller, E. H. (**2007**). "Vocal signalling in male southern elephant seals is honest but imprecise," Anim. Behav. **73**, 287–299.
Sayigh, L. S., Tyack, P. L., Wells, R. S., Solow, A. R., Scott, M. D., and Irvine, A. B. (**1999**). "Individual recognition in wild bottlenose dolphins: A field test using playback experiments," Anim. Behav. **57**, 41–50.
Schaller, G. B., Hu, J., Pan, W., and Zhu, J. (**1985**). *The Giant Pandas of Wolong* (University of Chicago Press, Chicago, IL).
Searby, A., and Jouventin, P. (**2003**). "Mother-lamb acoustic recognition in sheep: A frequency coding," Proc. Biol. Sci. **270**, 1765–1771.
Semple, S. (**2001**). "Individuality and male discrimination of female copulation calls in the yellow baboon," Anim. Behav. **61**, 1023–1028.

Sokal, R. R., and Rohlf, F. J. (**1995**). *Biometry: The Principles and Practice of Statistics in Biological Research* (W. H. Freeman and Co., New York).
Soltis, J., Leong, K., and Savage, A. (**2005a**). "African elephant vocal communication II: Rumble variation reflects the individual identity and emotional state of callers," Anim. Behav. **70**, 589–599.
Swaisgood, R. R. (**2004**). "Chemical communication in giant pandas," in *Giant Pandas: Biology and Conservation*, edited by D. G. Lindburg and K. Baragona (University of California Press, Berkeley, CA), pp. 106–120.
Swaisgood, R. R., Lindburg, D. G., and Zhou, X. P. (**1999**). "Giant pandas discriminate individual differences in conspecific scent," Anim. Behav. **57**, 1045–1053.
Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ).
Torriani, M. V., Vannoni, E., and McElligott, A. G. (**2006**). "Mother-young recognition in an ungulate hider species: A unidirectional process," Am. Nat. **168**, 412–420.
Tripovich, J. S., Charrier, I., Rogers, T. L., Canfield, R., and Arnould, J. P. Y. (**2008**). "Acoustic features involved in the neighbour-stranger vocal recognition process in male Australian fur seals," Behav. Processes **79**, 74–80.
Vannoni, E., and McElligott, A. G. (**2006**). "Individual acoustic variation in fallow deer (*Dama dama*) common and harsh groans: A source-filter theory perspective," Ethology **113**, 1–12.
Wiley, R. H., and Richards, D. G. (**1978**). "Physical constraints on acoustic communication in atmosphere—Implications for evolution of animal vocalizations," Behav. Ecol. Sociobiol. **3**, 69–94.
Xie, Z., and Gipps, J. (**2003**). *The 2003 International Studbook for the Giant Panda (Ailuropoda melanoleuca)* (Chinese Association of Zoological Gardens, Beijing, China).
Yin, S., and McCowan, B. (**2004**). "Barking in domestic dogs: Context specificity and individual identification," Anim. Behav. **68**, 343–355.
Zimmermann, E., and Hafen, T. G. (**2001**). "Colony specificity in a social call of mouse lemurs (*Microcebus ssp.*)," Am. J. Primatol **54**, 129–141.
Zimmerman, E., and Lerch, C. (**1993**). "The complex acoustic design of an advertisement call in male mouse lemurs (*Microcebus murinus*) and sources of its variation," Ethology **93**, 211–224.

Optical tracking of acoustic radiation force impulse-induced dynamics in a tissue-mimicking phantom

Richard R. Bouchard,[a] Mark L. Palmeri, Gianmarco F. Pinton, and Gregg E. Trahey
Department of Biomedical Engineering, Duke University, Box 90281, Durham, North Carolina 27708

Jason E. Streeter and Paul A. Dayton
Joint Department of Biomedical Engineering, University of North Carolina at Chapel Hill and North Carolina State University, Box 7575, Chapel Hill, North Carolina 27599

(Received 22 December 2008; revised 20 August 2009; accepted 26 August 2009)

Optical tracking was utilized to investigate the acoustic radiation force impulse (ARFI)-induced response, generated by a 5-MHz piston transducer, in a translucent tissue-mimicking phantom. Suspended 10-μm microspheres were tracked axially and laterally at multiple locations throughout the field of view of an optical microscope with 0.5-μm displacement resolution, in both dimensions, and at frame rates of up to 36 kHz. Induced dynamics were successfully captured before, during, and after the ARFI excitation at depths of up to 4.8 mm from the phantom's proximal boundary. Results are presented for tracked axial and lateral displacements resulting from on-axis and off-axis (i.e., shear wave) acquisitions; these results are compared to matched finite element method modeling and independent ultrasonically based empirical results and yielded reasonable agreement in most cases. A shear wave reflection, generated by the proximal boundary, consistently produced an artifact in tracked displacement data later in time (i.e., after the initial ARFI-induced displacement peak). This tracking method provides high-frame-rate, two-dimensional tracking data and thus could prove useful in the investigation of complex ARFI-induced dynamics in controlled experimental settings. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3238235]

PACS number(s): 43.80.Jz, 43.25.Qp, 43.25.Zx [TDM] Pages: 2733–2745

I. INTRODUCTION

Tissue elasticity imaging is a growing field[1–12] in medical diagnostic imaging whereby underlying mechanical properties of a tissue are gleaned from its mechanical response to an applied force. This response is generally tracked with ultrasound-based displacement estimators, which tend to be unidimensional and often suffer from undesirable sampling limitations. With the imaging of heterogeneous tissues with increasingly complex boundary conditions,[12–15] this induced dynamic response will likewise become more complex (e.g., shear wave anisotropy and plate wave propagation). Consequently, a tracking option with greater spatial and temporal sampling ability could offer valuable insight into these complex dynamics. We propose a novel tracking method that utilizes optical tracking, which is able to track with a high frame rate in two dimensions and is not corrupted by previously transmitted acoustic pulses. These optical tracking data, which were obtained in a tissue-mimicking phantom, are compared to finite element method (FEM) modeling results to further corroborate previously validated simulation results as well as offer corroboration to model elements currently without experimental validation (i.e., lateral displacement values and dynamics during a radiation force excitation).[16] Given the relative closeness (3.3–4.8 mm) of the phantom's most proximal boundary to the region of interest (ROI), this research is limited to transient effects (within the first 4 ms) that occur before this boundary is able to introduce artifacts to the ROI by way of shear wave propagation. Although penetration depth limitations will likely preclude this optically based method from becoming a clinically viable technique, it could still prove useful in understanding complex acoustic radiation force impulse (ARFI)-induced dynamics (e.g., at interface boundaries) or in further validating current FEM modeling results.

A. Acoustic radiation force

ARF is generated whenever an acoustic wave is either absorbed or scattered by an obstructing object. When this wave and object are assumed planar, this force is proportional to the average energy density of the incident wave and can be expressed as

$$\boldsymbol{F}_{\mathrm{rad}} = A\bar{E}\boldsymbol{d}_r, \tag{1}$$

where A equals the projected area on the object, $\bar{E}$ equals the temporal-average energy density of the incident wave, and $\boldsymbol{d}_r$ represents a drag coefficient.[9,17] This drag coefficient has two orthogonal components: the real component represents the force contribution in the direction of incident wave propagation and the imaginary component represents the contribution in the transverse direction. The drag coefficient is defined per unit energy density as

$$\boldsymbol{d}_r = A^{-1}\left(\Pi_a + \Pi_s - \int \gamma \cos\,\theta dA\right) - iA^{-1}\int \gamma \sin\,\theta dA, \tag{2}$$

[a]Author to whom correspondence should be addressed. Electronic mail: rrb@duke.edu

where Π_a and Π_s are the total powers absorbed and scattered, respectively, and γ is the scattered intensity, all per unit incident intensity, while θ is the scattering angle. In the case of many soft tissues and tissue-mimicking phantoms, where the predominant attenuation mechanism is absorption,[18] Eq. (1), which now represents a body force in the direction of incident wave propagation, can be simplified to

$$\boldsymbol{F}_{\mathrm{rad}} = \frac{2\alpha \boldsymbol{I}_{\mathrm{ta}}}{c}, \tag{3}$$

where α equals the absorption coefficient of the medium, $\boldsymbol{I}_{\mathrm{ta}}$ equals the temporal-average beam intensity, and c equals the speed of sound through the medium.[19,20]

When ARF is generated in an elastic solid, displacement is induced within the region of ultrasonic beam propagation; the initial magnitude of this displacement depends on the spatial variation in attenuation and intensity [as indicated by Eq. (3)] throughout this region. Along with this localized displacement, a dynamic ARF excitation produces shear waves, which travel in the transverse direction.[6] In a linear, isotropic, elastic medium, the speed of these shear waves can be expressed as

$$c_T = \sqrt{\frac{\mu}{\rho}} = \sqrt{\frac{E}{2(1+\nu)\rho}}, \tag{4}$$

where μ equals the shear modulus, ρ equals the density of the propagation medium, E is the Young's modulus, and ν is the Poisson's ratio.[21] Typical shear wave velocities in soft tissue range from 1 to 5 m/s.[22]

B. ARF-based tissue elasticity imaging

In the field of tissue elasticity imaging, induced tissue motion resulting from an applied force is analyzed. This force can be static or dynamic and can be applied externally or internally to the human body.[23] The research contained herein is concerned with the specific case when a dynamic force, of relatively short duration (<1 ms), is applied to (or just adjacent to) an internal ROI. An impulsive ARF, or "ARFI" as it will be referred to, is an effective way to create this transient, internal mechanical excitation. Significant research has been conducted on the use of an ARFI excitation for the purpose of tissue elasticity imaging. Current research initiatives have tended to focus on this transient response at the location of the ARFI application ("on-axis")[1–4] or at a lateral location outside of the excitation volume ("off-axis").[5–7] For the purpose of tissue characterization, the magnitude of the on-axis displacement is often inversely proportional to a tissue's mechanical stiffness[24] while the phase velocity of the transversely traveling wave (i.e., shear wave) created by the off-axis displacement is reflective of a tissue's shear modulus [Eq. (4)].[5,9]

C. Current tracking of ARF-induced dynamics

1. Ultrasonically based techniques

Displacement tracking of ARF-induced dynamics is typically achieved through the application of time-delay estimators on radiofrequency (RF) pulse-echo data. Pulse-echo, or "tracking," pulses are transmitted successively at a single lateral beam position to obtain displacement estimates through time for that location; this scheme is then translated laterally to obtain a two-dimensional field of view (FOV). Although this conventional technique can yield sub-micron displacement estimates in the axial dimension,[25] it is hindered by three fundamental limitations: poor tracking resolution in the lateral dimension, sampling limitations due to interference from previous pulses, and a large effective tracking kernel. First, conventional ultrasonically based lateral tracking yields a displacement estimate variance that is 40(focal distance/aperture width)^2 times worse than that achieved through axial tracking.[26] Second, interference from previous pulses limits both tracking pulse repetition frequency (PRF) and the ability to track during an ARFI excitation. Typically, echo signal from previous pulses (tracking or ARF excitation pulses) must be sufficiently attenuated before another pulse-echo tracking scheme can be initiated. Maximum PRF is limited by the desired depth of field, the duration of the excitation pulse (which can generally be ignored for pulse-echo excitations), and the medium's attenuation and speed of sound. If a point located at a focal depth of 3.8 cm was to be tracked in a tissue-mimicking phantom (i.e., 1540 m/s speed of sound), the highest possible PRF would be just over 20 kHz; this would be assuming no residual, interfering echo signal from propagation deep to this focus, which is unrealistic. In practice, PRFs are usually lower than the example given. In the case of tracking an ARFI-induced response, effects of a previously transmitted ARFI excitation tend to persist even longer, given the increase in pulse length (typically one to two orders of magnitude) over conventional pulse-echo excitations.[27] Consequently, pulse-echo tracking during and immediately following an ARFI excitation is generally not possible. Maleke *et al.*[28] were able to track tissue displacement during a continuous wave, amplitude-modulated ARF excitation for the purpose of harmonic motion imaging. By using a tracking pulse transmit frequency between harmonics of the ARF excitation transmit frequency, they were able to suppress interference from the ARF excitation through the application of a bandpass filter. This technique, however, has not been demonstrated with impulsive excitations. Third, the effective tracking kernel for an ultrasonically based method is disproportionally large in the lateral and elevation dimensions due to the inherent width ($\geq$hundreds of microns) of a beam's point spread function. If the magnitude of induced axial displacement varies throughout these lateral/elevational extents, the estimated displacement will tend to average this profile, which will result in an underestimation of the true peak displacement.[29]

2. Optically based techniques

Although there have been multiple nonultrasonically based tracking techniques employed for the purpose of elasticity imaging,[22,30] few have been able to track the transient dynamics generated from an ARFI excitation. Andreev *et al.*[31] first used a laser to track the ARF-induced shear wave dynamics of a 60–100-μm microsphere embedded in an elastic medium. This technique aligned the focus of a laser

on the microsphere, which was used as a shutter to occlude an opposing, coaxial photodiode. An ARFI excitation from an ultrasonic radiator, mounted in the transverse axis, produced a shear wave in the medium that caused the microsphere to move; the degree of photodiode obstruction was proportional to the microsphere's axial displacement. This technique is able to operate at a substantial depth (10 mm from the proximal boundary) and has good displacement estimation resolution (micrometer-order); yet, it only tracks a single particle in a single, orthogonal dimension and gives an indirect measurement of displacement within a range that is dependent on the radius of the obstructing particle. Bossy *et al.*[32] similarly employed the use of a laser, but they instead measured the decorrelation (within 272×272-μm^2 kernels) of received optical speckle patterns, resulting from transmission through a translucent phantom, to gain an indirect measure of internal phantom dynamics generated by shear wave propagation; no embedded microspheres were necessary with this method. This technique is able to function at a significant depth (20 mm) and provides information regarding optical and shear mechanical properties of the tissue, but it offers limited resolution and frame-rate capabilities (millimeter-order and 2 kHz, respectively), does not provide a direct measure of local displacement, and is unable to discern axial from lateral motion.

Perhaps the most standard use for optical tracking of ARF-induced dynamics has been in the investigation of the force's effect on microbubbles, a common ultrasound contrast agent. Using a microscope and attached camera, Dayton *et al.*[33] were able to observe the behavior of a microbubble aggregate when exposed to an ARF excitation. This work was further expanded by Dayton *et al.*[35] and Palanchon *et al.*[34] when they continued to investigate, with a high-speed camera, the effects of ARF on a single microbubble. The experimental configuration presented in this paper is roughly modeled after the basic setup often employed in microbubble experimentation. Bouchard *et al.*[36] recently utilized a similar experimental configuration to track, in two dimensions, the ARF-induced dynamics on the surface of a tissue-mimicking phantom and catheter-based device; these tracking data were then validated with a conventional ultrasonically based tracking technique. Although Bouchard *et al.* were able to track in two dimensions with sub-micron resolution throughout the complete dynamic response, their experimental setup was strictly limited to superficial tracking. Additionally, they included an extra transducer for ultrasound-based tracking, which was not incorporated into this study.

If one were to employ a translucent medium in an optical tracking study, investigation of induced dynamics would not be limited to superficial regions. In the case when light is transmitted through an opaque medium, its intensity decays in an exponential manner which is characterized by the Lambert–Beer law.[37,38] This decay, which tends to be scattering dominated, increases with increased scatterer density in the medium and increased transmission distance of a photon through the medium. Scattering can also cause a defocusing effect to occur when a region within this opaque medium is viewed through an optical microscope. As the number of scatterers between the objective and the focal plane increases, there is an increased contribution from multiply scattered and off-axis photons, which contributes to a loss of focus and contrast in excessively thick samples.

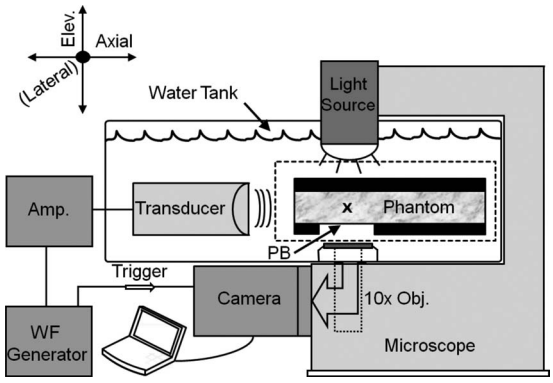


FIG. 1. Experimental setup overview. Dotted box outlines the phantom setup, which is detailed in Fig. 2. "PB" denotes the relative location of the phantom's proximal boundary. "X" indicates transducer/microscope foci. Setup equipment and spacings are not drawn to scale. Presented axes are same as those referenced throughout the paper.

II. METHODS

A. Optical phantom construction

A gelatin-based, translucent, tissue-mimicking phantom was constructed. The gel solution recipe, which was modeled after a formulation by Takegami *et al.*,[39] consisted of the following ingredients (with mass percents given in parentheses): 100 Bloom type-A gelatin (6.8%; Vyse Gelatin Co., Schiller Park, IL), *n*-propanol (3.7%), egg white (35.6%), de-ionized water (53.4%), and 25%-glutaraldehdye (0.5%). Eighteen drops of 10-μm black polystyrene microspheres (Polysciences Inc., Warrington, PA) were added to serve as optical tracking markers. Egg white was included to increase the phantom's acoustic absorption (when compared to gelatin alone) without drastically increasing the optical scattering of the medium. Phantom production was achieved through a similar protocol as that outlined by Hall *et al.*[40] The phantom was initially cast in a 10 cm (diameter) $\times$ 2.5 cm (height) cylindrical mold with acoustic windows on both ends to allow for thru-transmission measurements, which yielded speed of sound and acoustic attenuation measurements (using a substitution method[41]) of 1570 m/s and 0.7 dB/cm (at 5 MHz), respectively, for the experimental phantom. The phantom [Fig. 2(a)] was then removed from the mold, sliced to a 15-mm height, and cut along a 5-cm chord to produce a flat surface along the circumferential perimeter.

B. Experimental setup

All experiments were conducted on the equipment setup depicted in Fig. 1. The basic configuration consisted of a microscope and attached high-speed camera with an ARF-generating ultrasound transducer mounted in the transverse axis. The foci of the microscope and transducer were nearly coincident and positioned in the optical phantom, which was supported by a custom acrylic holder in a water tank.

The ultrasound, or "push," transducer (IL0506HP, Valpey Fisher Corp., Hopkinton, MA) is a single-element

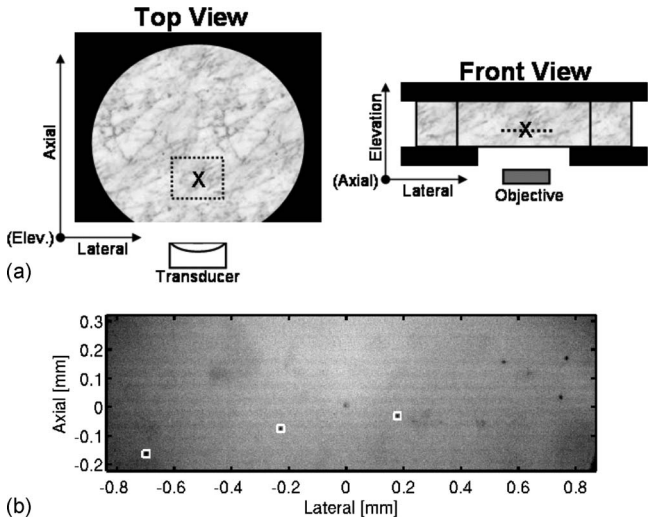


FIG. 2. Phantom setup (a) and microscope FOV example (b). In the right and left phantom setup diagrams, front (side nearest transducer face) and top (upper acrylic plate removed) phantom views are given, respectively. Phantom is denoted by marble shading; acrylic holder, by black shading. Relative positions of the microscope objective and push transducer are identified. Dotted line and box indicate imaging plane and FOV, respectively. "X" indicates approximate transducer/microscope foci. Note cutout in acrylic holder base to allow for optimum microscope visualization. In FOV example diagram, an experimentally utilized (Exp. 6) FOV screen capture (pre-excitation) is given. Six "trackable" microspheres are visible—three in kernel boxes and three (black dots, lateral positions 0.5–0.8 mm) lacking kernel boxes.

piston transducer with a 38-mm focus, 19-mm diameter, 5-MHz center frequency, and a full width at half-maximum (FWHM) beamwidth of 650 μm at the focus (690 μm at 40 mm). The push transducer was mounted to an *XYZ* micropositioning stage (Edmund Optics, Barrington, NJ), with 100-μm precision, for the purpose of precision adjustment. An Olympus IX71 microscope (Olympus America Inc., Melville, NY) fitted with a 10× objective and opposing, coaxial, 100-W halogen illumination source (U-LH100L-3, Olympus America Inc., Melville, NY) was used to gain a magnified visualization of the imaging plane. A high-speed camera (Fastcam SA1, Photron USA Inc., San Diego, CA) was coupled to the microscope to allow for digital frame capture.

To ensure that the foci of the push transducer and microscope were approximately coincident, a needle hydrophone (HNC-0400, Onda Corp., Sunnyvale, CA) with a 400-μm tip diameter was first centered (axially and laterally) within the microscope's FOV and placed in focus (elevationally). With the hydrophone fixed, the push transducer was then adjusted with the micro-positioning stage until the peak amplitude of a repeated 1-cycle burst was placed at the approximate center of the hydrophone in the lateral and elevation dimensions. Due to mechanical limitations on the micropositioning stage, it was not possible to place the spatial intensity peak of the transducer's output at the center of the microscope's FOV in the axial dimension. Instead, a point approximately 2 mm deep to the absolute axial focus was placed at the center of the microscope's FOV. The optical phantom's proximal boundary, which was partially supported by a transparent acrylic holder attached to a second micropositioning stage, was then placed at the coincident transducer/microscope foci. Using the vertical micropositioning stage adjustment, depth into the phantom (relative to the microscope focus) could then be determined based on elevational translation. The flat surface along the circumferential perimeter of the phantom was then positioned 13 mm from the transducer face; this 13-mm water standoff was maintained for all experiments. The phantom was then translated laterally until a reasonably homogeneous region with in-plane microspheres was positioned within the microscope's FOV. The same phantom region was not used throughout experimentation; all utilized phantom regions, however, were within a few millimeters laterally of one another. The push transducer was driven by an arbitrary waveform generator (AWG2021, Tektronix Inc., Wilsonville, OR) and amplified by 55 dB using a RF amplifier (3200LA, Electronic Navigation Industries, Rochester, NY). The ARFI excitation used in this study was measured with a PVDF membrane hydrophone (804–107, Sonic Technologies, Hatboro, PA) to have an *in-situ* intensity ($I_{SPPA.15}$) of 2.5 kW/cm^2 and a mechanical index of 1.8. The output of the waveform generator was synchronized with the high-speed camera to ensure that the video acquisition trigger was coincident with the initiation of ARFI excitation generation. It is important to note that approximately 25 μs of pulse propagation time followed the trigger signal before an ARFI pulse *started* being absorbed at the ROI.

TABLE I. Experimental parameters.

Experiment No.	1	2	3	4	5	6	7
Pulse length (ms)	0.1	0.1	0.2	0.2	0.2	0.4	0.4
Frame rate (kHz)	6.25	10	10	5	36	24	24
Depth (mm)	4.8	4.8	4.8	3.7	3.3	3.3	3.3
FOV offset (mm)	0	0	0	0	0	0	−2:3[a]
No. of trials	1	1	1	3	3	3	1

[a]FOV shifted to six laterally offset locations: 1, 2, 3, −1, −2, and 0 (on-axis) mm; one experimental trial was conducted at each.

C. Experimental protocol

Each experiment was conducted in a similar manner. First, an ARFI excitation, which consisted of a single 5-MHz tone burst of a particular pulse length (given in Table I), was transmitted into the optical phantom at a specific depth from the phantom's proximal boundary. As the push transducer began transmitting the ARFI excitation, the video camera was triggered to capture digital images of the microscope's FOV at a frame rate specified in Table I (the shutter speed was set to the inverse of the frame rate); 8–10 pre-excitation frames were also captured to obtain information regarding the phantom's initial condition. Depending on the experiment, the microscope's FOV was either centered about the approximate location of the push transducer's lateral focus (i.e., on-axis) or laterally offset (but maintaining the same axial/elevation positions) by a specified amount (i.e., off-axis). Specific parameters for each experiment are listed in Table I and depicted in Fig. 3; experiment numbers reflect the actual experimental order. It is important to note that Exp. 8 (presented in Sec. II F) did not involve optical track-

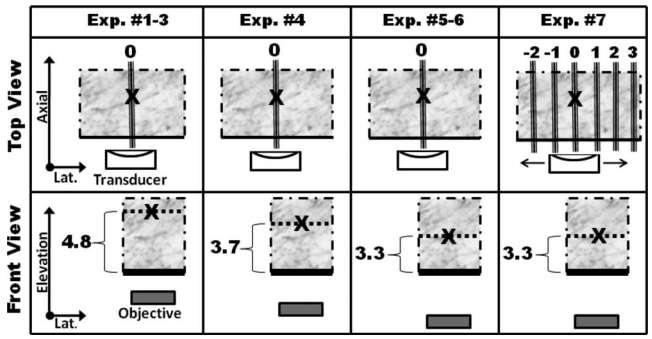


FIG. 3. Transducer/objective positions for all optical experiments. Figure elements (e.g., transducer and objective) and orientations correspond to those depicted above in Fig. 2. Top view provides lateral position of transducer excitation beam (denoted by three vertical lines); front view provides position of microscope focus relative to proximal phantom boundary (denoted by a thick solid line). "X" indicates microscope focus. Entire phantom is not depicted in each figure; only phantom regions nearest transducer/objective face shown. Lateral offsets/depths are given in millimeters. Axes have third dimension label omitted due to space considerations.

ing and thus in not included in the summary table or figure. Camera frame rates and focus depths were limited to 36 kHz and 4.8 mm, respectively, based on approximate points at which light intensity and image contrast became too poor in experimental data to perform reliable tracking. Experimental acquisitions were separated by a minimum of approximately 1 min, during which time the halogen illumination source was turned off, to mitigate phantom heating. Using a digital temperature probe (51-II, Fluke Corporation, Everett, WA), a peak temperature increase of 0.4 °C was measured over 10 s in the phantom due to the illumination source.

The number of trials for each experiment indicates the number of times the experiment was repeated. In the case of Exp. 7, the experiment was conducted once at each FOV offset. To effectively change the microscope's FOV relative to the push transducer's focus, the microscope and phantom were kept fixed while the transducer was precision-translated laterally (depicted in the upper, right-most image in Fig. 3). The transducer was first translated to the five off-axis locations before finally being returned to its approximate original position.

D. Data processing

Displacement tracking was achieved by manually selecting the centers of all in-focus microspheres (typically 1–6) within the microscope's FOV [Fig. 2(b)] and automatically enclosing each in 24×24-μm^2 tracking kernels. Tracking kernels were then translated within a limited radius at each time step, and a correlation coefficient (CCoef), relative to a reference kernel taken from a pre-excitation frame, was calculated for each translation step; the translation corresponding with the highest CCoef for each frame was deemed the most accurate displacement estimate. The post-magnification pixel separation for digitized images was measured with a calibrated reticle slide to be 2 μm; two-dimensional linear interpolation was utilized to allow for sub-sample shifts of 0.5 μm, a reasonable limit given the average signal-to-noise ratio (SNR) encountered in the video data during experimentation.

Shear wave and lateral displacement wave velocity estimates were obtained by implementing a time-of-flight method, the lateral time-to-peak (TTP) algorithm,[42] on displacement data acquired from kernels outside of the ARFI excitation volume (i.e., off-axis). In general, the time-to-peak displacement (TTPD) for a specific kernel was plotted as a function of its lateral distance from the excitation center. A linear regression fit was then performed on these data to yield a wave velocity estimate (obtained from the fit's inverse slope). The square of a fit's correlation coefficient was calculated to indicate its degree of linearity, with $r^2=1$ indicating perfectly linear data. When TTPD values were calculated for lateral displacement data, a running average was applied through time with a 167-μs kernel length. The same filter was applied to axial displacement data prior to differentiating for the purpose of obtaining kernel velocity estimates in Exp. 5.

In an effort to concisely quantify the tracking quality of displacement data, three metrics related to the CCoef are presented in Sec. III with all tracked data: lowest (among all kernels represented in the figure) average CCoef, highest standard deviation, and lowest CCoef value achieved—all through time for a single kernel but not necessarily the same kernel. These parameters, which are listed in an array ($CCoef_{worst}=[\min(avg),\max(std),\min(value)]$), are meant to give a worst case depiction. Typical per-frame CCoef values were above 0.85 for Exps. 1–5 and above 0.9 for Exps. 6 and 7; CCoef values were always lowest during transmission of the ARFI excitation.

E. FEM modeling

Three-dimensional FEM models of the dynamic response of elastic media to ARFI excitations were used to confirm the underlying physical mechanisms responsible for the experimentally observed dynamics. These models utilized a mesh of 1,175,000 eight-noded, linear cubic elements with 0.1-mm node spacing. The model was performed in three dimensions using quarter-symmetry (about the transducer's axis of symmetry) boundary conditions. The lateral/elevation dimensions of the mesh extended 5 mm from this axis of symmetry while the axial dimension extended 6 cm from the transducer face. Degrees of freedom for the symmetry faces were set for their appropriate symmetry conditions, the "outer" faces were unconstrained, and the bottom and top boundaries were fully constrained. The material was modeled as a linear, isotropic, elastic (i.e., no viscosity) solid with $E=3.4$ kPa, $\nu=0.499$, and $\rho=1.0$ g/cm^3. The Young's modulus value was determined through shear wave velocity data obtained in Exp. 8 (presented in Sec. II F) and Eq. (4), a relationship which was experimentally validated in gelatin-based phantoms.[16]

Simulation of the acoustic intensity associated with the piston utilized in these experiments was performed using FIELD II, a linear acoustics modeling software package.[43,44] The piston (38-mm focus, 19-mm diameter) was simulated in FIELD II using the xdc_concave function with 0.5-mm square mathematical sub-elements. No attenuation was simulated in the water path (0–13 mm), and the effects of nonlinear wave

propagation in the water and reflections at the water/phantom interface were not taken into account in these simulations.

The radiation force field was applied as point loads to individual nodes in the region of excitation. The magnitude of these point loads was determined using Eq. (3), where the intensity was averaged over the adjacent element volume for a given node, and the attenuation was 0.15 dB/cm MHz. Forces were applied for the specified pulse length duration and directed along the Poynting vector as a function of position in the region of excitation. For the on-axis result, simulation parameters were based on Exp. 5; for the off-axis result, they were based on Exp. 7.

The dynamic response of the elastic solid was solved through the balance of linear momentum and using LS-DYNA3D (Livermore Software Technology Corp., Livermore, CA), which implemented an explicit, time-domain, integration method. Single-point quadrature was used with Flanagan–Belytschko integration stiffness form hourglass control to avoid element locking and to reduce numerical artifacts. Two-dimensional linear interpolation was employed on the modeling result to obtain displacement values at the specific lateral/axial kernel positions. To obtain registration between model results and optically tracked data, the experimental FOV center was assumed to be perfectly centered in the lateral/elevation beam dimensions and located 40 mm (i.e., 2 mm deep to the simulated axial focus) from the transducer face in the axial dimension. Pulse propagation time to the ROI was accounted for in the model results. All of these modeling methods have been applied previously to gelatin-based phantoms, as detailed by Palmeri *et al.*[16]

F. Ultrasonically based shear wave velocimetry

A SONOLINE Antares™ ultrasound system with a VF10-5 commercial linear array (Siemens Medical Systems, Ultrasound Group, Issaquah, WA) was used for independent ultrasonically based validation of the shear wave velocity estimate in the phantom. ARFI excitations were generated from and tracked with the linear array probe. Inclusion of egg white solution in the gelatin-based phantom generated enough backscatter to allow for reliable ultrasonically based tracking. From the tracked data, shear wave velocity estimates were obtained with the lateral TTP algorithm, the implementation of which is described in detail by Palmeri *et al.*[42] Shear wave velocity estimates were acquired at three independent regions in the experimental phantom. This protocol will be referred to as Exp. 8.

III. RESULTS

Optical tracking results, which are presented first, focus on three specific aspects: on-axis dynamics (Exps. 1–6), off-axis dynamics (Exp. 7), and proximal boundary effects (Exps. 1–4 and 6). These results are then compared to matched FEM modeling and ultrasonically based shear wave velocimetry (Exp. 8) results.

A. On-axis dynamics

Figure 4 depicts the on-axis dynamic response for Exp. 6 from Table I and Fig. 3. The dynamic responses of six embedded microspheres (each tracked with a single kernel) are presented by identifiably colored traces. Specifics regarding kernel location (relative to the bottom image in Fig. 2), displacement peaks (positive or negative), and TTPD values are presented in Table II.

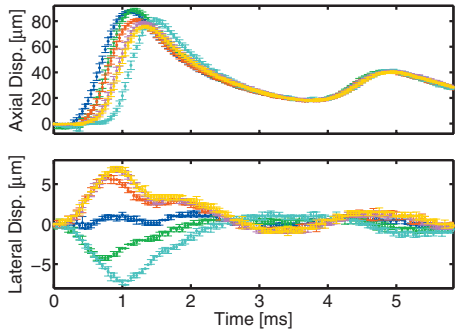


FIG. 4. Tracked axial (above) and lateral (below) displacement induced by 0.4-ms ARFI excitation. Each trace depicts the mean and standard deviation ($N=3$) with dots and error bars, respectively, for a specific tracked kernel at sampled time points (24-kHz frame rate). Corresponding trace colors (blue, green, red, cyan, magenta, and yellow —represented by first letter of respective color name) and coordinates of plotted kernels ($CCoef_{worst}=[0.85,0.05,0.65]$) are presented in Table II. Coordinates for all figures are relative to the FOV center. Time scale commencement for all figures is coincident with the initiation of ARFI excitation transmission. Unless otherwise noted, abscissa and ordinate scales are the same for plots within a given figure and all plotted data are unfiltered.

In the axial displacement plot (above), there are two displacement peaks: one resulting from the initial ARFI excitation (occurring between 1.13–1.46 ms) and one resulting from the proximal boundary shear wave (PBSW) reflection (occurring around 5 ms). In the lateral displacement plot (below), the direction of a kernel's displacement depends on its lateral position relative to the excitation center; a kernel will tend to displace away from the excitation's central axis. Much like the axial displacement plot, there are two peaks in absolute displacement presented in the lateral displacement plot: one resulting from the initial ARFI excitation (occurring between 0.71–1.00 ms) and a subtle one resulting from the PBSW reflection (occurring around 5 ms).

The peak displacement and TTPD values for each kernel are listed in Table II. Peak lateral displacement magnitude and TTPD, in both the axial and lateral dimensions, values tend to increase as a kernel's lateral distance from the excitation center increases. A notable exception to this trend is the cyan trace, which peaked (in axial and lateral displacement) latest and experienced the greatest lateral displacement despite having a reported absolute lateral distance (0.70 mm)

TABLE II. Exp. 6 on-axis (Fig. 4) kernel parameters.

Trace color	B	G	R	C	M	Y
Axial pos. (mm)	−0.04	−0.08	0.15	−0.17	0.03	0.17
Lateral pos. (mm)	0.18	−0.23	0.55	−0.70	0.75	0.77
Peak ax. disp. (μm)	87.9	88.9	81.0	80.5	78.6	75.9
Axial TTPD (ms)	1.13	1.17	1.21	1.46	1.29	1.33
Peak lat. disp. (μm)	0.9[a]	−4.2	5.6	−7.2	6.6	6.8
Lateral TTPD (ms)	0.96[a]	0.71	0.79	1.00	0.92	0.92

[a]Search region limited to first 1.25 ms.

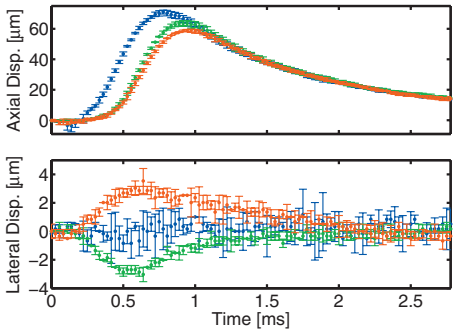


FIG. 5. Tracked (36 kHz) axial (above) and lateral (below) displacement induced by 0.2-ms ARFI excitation. Corresponding trace colors and coordinates of plotted kernels ($CCoef_{worst}$=[0.73,0.06,0.46]) are presented in Table III. Both plots are truncated to allow for better visualization of early-time dynamics; windowing removed PBSW reflection peak (occurring around 5 ms). All other plotting conventions are the same as those used in Fig. 4.

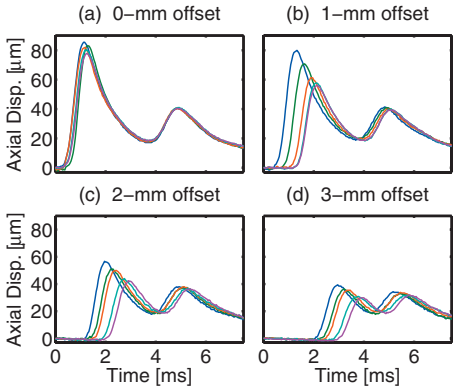


FIG. 6. Tracked (24 kHz) axial displacement resulting from 0.4-ms ARFI excitation centered relative to FOV of on-axis plot (a). FOVs for the other three plots [(b)–(d)] are increasingly offset laterally (1, 2, and 3 mm, respectively). Each trace represents a single tracked kernel ($CCoef_{worst}$=[0.86,0.03,0.76]; same for Fig. 7). Corresponding trace colors and kernel coordinates are presented in Table IV.

that was closer than the magenta and yellow kernels (with 0.75 and 0.77 mm lateral distances, respectively).

Figure 5 depicts the on-axis dynamic response for Exp. 5. The dynamic responses of three embedded microspheres are represented by identifiably colored traces. Similar kernel parameters to those offered for Exp. 6, with the notable inclusion of peak axial velocities and corresponding occurrence times, are presented in Table III.

As in the previous on-axis case, peak lateral displacement and axial/lateral TTPD values increase while peak axial displacement values decrease as a kernel's absolute lateral offset increases. A kernel's peak axial velocity and the corresponding time at which it occurs likewise increase as absolute lateral distance increases. In the axial displacement data (above), the blue kernel experienced a significant decrease in its CCoef, to 0.46, during absorption of the ARFI excitation (around 0.15 ms). This resulted in noticeably increased tracking jitter in its axial displacement plot (blue trace) and what is assumed to be an artifactual, mean negative displacement.

B. Off-axis dynamics

Figures 6 and 7 summarize the off-axis response resulting from Exp. 7 (0–3 mm offsets). Figure 6 presents axial displacements, while Fig. 7 presents lateral displacements at four FOV offsets. Appropriate kernel parameters are presented in Table IV and apply to both figures. In all plots, two displacement peaks occur for each trace. The earlier peak is due to the initial excitation volume displacement or the outgoing shear wave it creates; the later peak is due to the PBSW reflection.

In the on-axis plots [0-mm offset— Figs. 6(a) and 7(b)], similar traits are observed when compared to the previous two on-axis experiments (Exps. 5 and 6). For the most part, kernels with greater lateral offset achieve less axial displacement but greater lateral displacement and axial/lateral TTPD values. Much like the first example presented (Exp. 6), there was one kernel, represented by the green trace, that had lateral displacement and TTPD values that were greater than expected. Additionally, the blue kernel experienced substantial negative lateral displacement (−1.7 μm) despite reportedly being located in the positive lateral region (0.17 mm).

In the off-axis, axial displacement plots [Figs. 6(b)–6(d)], observed behavior is similar to noted on-axis dynamics. Kernels further from the ARFI excitation center achieve a displacement peak that is less in magnitude and occurs later in time. As FOV offset increases, there is a greater amount of time following transmission of the ARFI excitation when no displacement is observed, indicating that the outgoing shear wave has not yet reached those more distal kernels. The PBSW reflection peak in each axial displace-

TABLE III. Exp. 5 on-axis (Fig. 5) kernel parameters.

Trace color	B	G	R
Axial pos. (mm)	−0.13	0.03	0.17
Lateral pos. (mm)	0.04	−0.20	0.50
Peak ax. disp. (μm)	71.0	63.6	58.9
Axial TTPD (ms)	**0.78**	**0.92**	**0.94**
Peak lat. disp. (μm)	−0.3[a]	−2.8	3.0
Lateral TTPD (ms)	**0.47**[a]	**0.56**	**0.61**
Peak ax. velocity (m/s)	0.184	0.174	0.154
Velocity peak time (ms)	0.47	0.61	0.61

[a]Search region limited to first 0.75 ms.

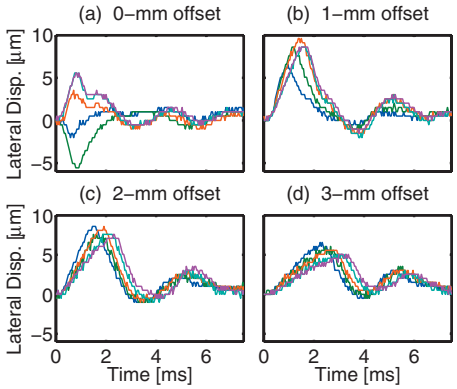


FIG. 7. Tracked (24 kHz) lateral displacement resulting from ARFI excitation centered relative to FOV of on-axis plot (a). Same kernels, FOVs, and plotting conventions are used as those presented in Fig. 6.

TABLE IV. Off-axis (Figs. 6 and 7) kernel parameters,

	(a) 0-mm offset				
Axial pos. (mm)	−0.07	−0.12	0.11	−0.01	0.13
Lateral pos. (mm)	0.17	−0.23	0.55	0.75	0.77
Peak ax. disp. (μm)	85.5	83.0	82.0	80.5	77.9
Axial TTPD (ms)	1.13	1.25	1.17	1.25	1.25
Peak lat. disp. (μm)	−1.7	−5.5	3.0	5.5	5.4
Lateral TTPD (ms)	0.63	0.83	0.71	0.83	0.83
	(b) 1-mm offset				
Axial pos. (mm)	−0.10	−0.05	0.13	0.01	0.15
Lateral pos. (mm)	0.77	1.17	1.55	1.75	1.77
Peak ax. disp. (μm)	79.9	70.8	61.2	57.7	56.2
Axial TTPD (ms)	**1.33**	**1.63**	1.88	**2.08**	2.13
Peak lat. disp. (μm)	6.2	8.4	9.4	8.6	8.6
Lateral TTPD (ms)	**0.92**	**1.17**	1.42	**1.58**	1.58
	(c) 2-mm offset				
Axial pos. (mm)	−0.10	−0.02	−0.06	0.13	0.01
Lateral pos. (mm)	1.77	2.00	2.17	2.54	2.74
Peak ax. disp. (μm)	56.7	51.1	50.1	44.0	42.5
Axial TTPD (ms)	1.96	2.21	**2.42**	2.75	**2.92**
Peak lat. disp. (μm)	8.5	7.3	8.2	7.6	7.3
Lateral TTPD (ms)	1.50	1.58	**1.83**	1.96	**2.33**
	(d) 3-mm offset				
Axial pos. (mm)	−0.10	−0.02	−0.06	0.13	0.01
Lateral pos. (mm)	2.76	2.99	3.17	3.54	3.74
Peak ax. disp. (μm)	39.5	36.4	35.4	30.9	29.9
Axial TTPD (ms)	2.96	3.21	3.33	3.75	3.88
Peak lat. disp. (μm)	6.3	5.7	5.6	4.9	5.1
Lateral TTPD (ms)	2.29	2.38	2.54	3.00	3.08
Trace color	B	G	R	C	M

ment plot occurs around 5 ms. Yet, these TTPD values progressively spread out and corresponding peak displacements attenuate as kernel offset increases. In the on-axis case (a), PBSW reflection peaks are virtually coincident while they are clearly dispersed, with a range of about 1 ms, and sequentially attenuated in the 3-mm offset case (d).

In the off-axis, lateral displacement plots [Figs. 7(b)–7(d)], observed dynamics are more complex than in the axial case. With all kernels in the positive lateral region, lateral displacements were consequently positive in all cases. Additionally, TTPD values increased as kernel offsets increased. Peak lateral displacement values, however, did not simply increase with increasing lateral kernel distance. In the 1-mm offset plot (b), peak displacement begins to increase with increasing kernel distance for the closest three kernels, but then it decreases for the furthest two kernels. The behavior in the 2-mm offset plot (c) is even less clear, with peak displacements tending to oscillate with increasing kernel offset. In the 3-mm offset plot (d), peak displacement tends to decrease with increasing kernel distance. Unlike the off-axis, axial displacement plot, there is no displacement delay with increasing FOV offset. With all three offsets, monotonically increasing lateral displacement is observed almost immediately after ARFI excitation transmission.

Table V summarizes all of the shear wave or lateral displacement wave velocities derived from presented data

TABLE V. Shear and lateral displacement wave velocities.

Source	Velocity (m/s)	r^2	Exp. No.
Optical axial	1.04 ± 0.02	0.99 ± 0.001	7[a]
FEM axial	1.04	1.00	7
Ultrasonically	1.07 ± 0.07	0.95 ± 0.02	8
Optical lateral	1.26 ± 0.25	0.98 ± 0.008	7[a]
FEM lateral	1.16	1.00	7
Boundary effect	0.97	0.99	1–4 and 6

[a]Only data from 2, 3, and −2 mm offsets used to ensure displacement estimates were not contained within the excitation volume.

sets. When possible, a mean and standard deviation (N=3) are offered. Shear wave speed estimates (i.e., wave speeds not derived from lateral displacements) yield good agreement, all with estimates around 1 m/s. The FEM modeling-derived ("FEM lateral") lateral displacement wave speed is significantly higher than any of the shear wave estimates, with a speed of 1.16 m/s. The mean optically derived ("optical lateral") lateral displacement wave speed is likewise higher than any mean shear wave speed. All wave velocity data, with mean $r^2 \geq 0.95$, are quite linear.

C. Proximal boundary effects

The effect of the proximal boundary (Exps. 1–4 and 6), which ranged from 3.3 to 4.8 mm from the imaging plane, was analyzed. Figure 8(a) presents a comparison of the time to peak PBSW reflection (axial) displacement for different proximal boundary distances. The plot illustrates that as the boundary gets progressively further away, the peak of the shear wave reflection it causes occurs later in time. This peak time and the propagation distance (which is approximately twice the boundary distance) share a direct, linear relationship, which is indicative of a constant shear wave velocity (0.97 m/s, Table V) in the medium. Figure 8(b) shows the dynamic response (axial displacement) of a single kernel, which was approximately 4.8 cm from the phantom's proximal boundary (Exp. 2). This plot demonstrates that the phantom completely recovers following the shear wave reflection artifact; an average of only 0.14 μm of residual displacement exists in the last 1.5 ms of plotted displacement.

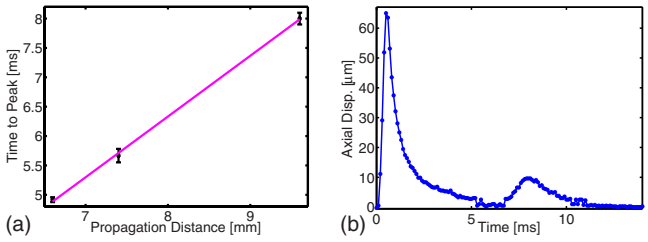


FIG. 8. (Color online) Proximal boundary effect comparison (a) and demonstration of phantom recovery (b). In the boundary effect plot, the mean and standard deviation (N=3) for time to the shear wave reflection peak for three proximal boundary distances [3.3 (Fig. 4), 3.7 ($CCoef_{worst}$ =[0.80,0.06,0.45]), and 4.8 mm ($CCoef_{worst}$=[0.85,0.03,0.58])] is plotted as a function of propagation distance (i.e., twice boundary distance). Linear regression fit (r^2=0.99) to means of these data is also plotted. In the recovery plot, the dynamic response of a single kernel is tracked until full recovery.

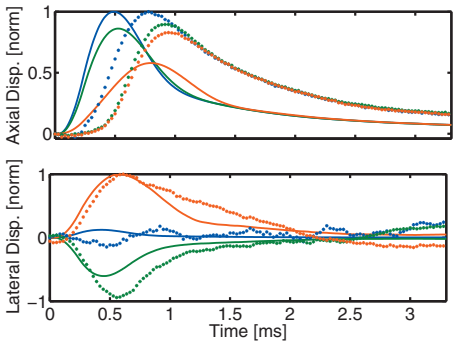


FIG. 9. Comparison of experimental (dots) and FEM (line) results for 0.2-ms, on-axis ARFI excitation. Axial (above) and lateral (below) displacement profiles are presented with normalized displacement scales. Trace coloring corresponds to kernel parameters listed in Table VI. Running average (167-μs kernel) was applied to all experimental lateral displacement data prior to plotting for FEM comparisons.

TABLE VI. On-axis FEM (Fig. 9) kernel parameters.

Experimental results[a]			
Peak ax. disp. (norm)	1.00	0.90	0.83
Peak lat. disp. (norm)	−0.15[b]	−0.94	1.00
FEM results			
Peak ax. disp. (norm)	1.00	0.86	0.58
Axial TTPD (ms)	0.49	0.53	0.80
Peak lat. disp. (norm)	0.12	−0.61	1.00
Lateral TTPD (ms)	0.43	0.45	0.58
Peak ax. velocity (m/s)	0.288	0.231	0.082
Velocity peak time (ms)	0.21	0.21	0.38
Trace color	B	G	R
Lateral pos. (mm)	0.04	−0.20	0.50
Axial pos. (mm)	−0.13	0.03	0.17

[a]TTPD values marked as bold in Table III.
[b]Search region limited to first 0.75 ms.

D. FEM modeling results

Figures 9 and 10 present a comparison of FEM modeling results and experimental tracking data. In the on-axis case, tracking data from Exp. 5 are compared to their matched FEM modeling result; in the off-axis case, FEM results are compared to data from Exp. 7 (1 and 2 mm offsets). Displacement traces are normalized to the peak value in their respective data set for all comparisons. For the regions and time frames analyzed, appreciable displacement was only observed in the simulation results through a depth range within ±1 cm of the transmit focus (i.e., 28–48 mm).

In the on-axis comparison (Fig. 9), there is reasonable agreement between peak normalized displacement values for both axial and lateral data. From the axial data (above), peak displacement and velocity decrease, while TTPD values increase with increasing kernel offset for both (i.e., FEM and experimental) data sets. Additionally, the leading edge of the displacement traces share a similar profile, which is consistent with peak velocities being of the same order and having percent differences all below 62%. There is a significant difference in TTPD values between data sets: displacement peaks in the simulated data occur much earlier than in the experimental data. If the blue traces are shifted to be coincident, however, percent differences between experimental and simulated TTPD values for the green and red traces are only 17% and 21%, respectively, if shifted by the same amount. Recovery profiles (i.e., portion of the trace following peak displacement) differ significantly between the two data sets, with the simulated data recovering noticeably faster than the experimental data.

The lateral displacement data (below in Fig. 9) share better agreement. Peak amplitudes and TTPD values increase with increasing kernel offset for both sets of results. Disregarding the nearly centered kernel (blue), which has displacement traces that are both quite flat, displacement is always directed away from the central axis in both tracking data. Much like the axial result, the leading edges of the displacement traces for the two more distal kernels exhibit a similar profile. Yet, unlike the axial result, the recovery profiles of these lateral data are fairly similar. The percent errors between TTPD values for the three kernels, in order of increasing kernel offset, are 9%, 22%, and 5%. On average for the three kernels, non-normalized, peak axial displacements are 142 times larger than non-normalized, peak lateral displacement magnitudes for the FEM data while they are only 29 times larger for the experimental data.

In the off-axis comparison (Fig. 10), agreement is quite good, particularly with the axial displacement data (above). Although recovery is still slower for the experimental data, the overall displacement profiles are quite similar between

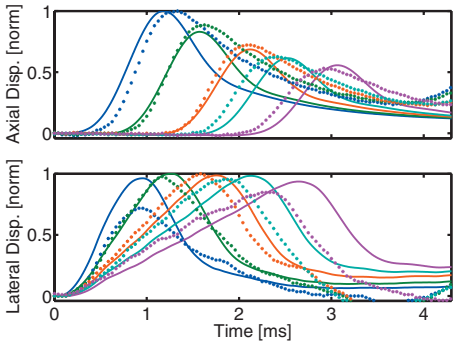


FIG. 10. Comparison of experimental (dots) and FEM (line) results for 0.4-ms, off-axis ARFI excitation. Axial (above) and lateral (below) displacement profiles presented with normalized displacement scales. Trace coloring corresponds to kernel parameters listed in Table VII.

TABLE VII. Off-axis FEM (Fig. 10) kernel parameters.

Experimental results[a]					
Peak ax. disp. (norm)	1.00	0.89	0.72	0.63	0.53
Peak lat. disp. (norm)	0.72	0.98	1.00	0.95	0.85
FEM results					
Peak ax. disp. (norm)	1.00	0.83	0.69	0.61	0.56
Axial TTPD (ms)	1.18	1.57	2.12	2.53	3.06
Peak lat. disp. (norm)	0.96	1.00	0.99	0.98	0.94
Lateral TTPD (ms)	0.95	1.27	1.76	2.13	2.65
Trace color	B	G	R	C	M
Lateral pos. (mm)	0.77	1.17	1.75	2.17	2.74
Axial pos. (mm)	−0.10	−0.05	0.01	−0.06	0.01

[a]TTPD values marked as bold in Table IV.

the two data sets. Unlike the on-axis data, experimental and FEM traces are relatively coincident, with TTPD value percent differences, in order of increasing kernel offset, of only 12%, 4%, 2%, 4%, and 5%. Normalized peak axial displacements shared similar agreement, with percent differences, in increasing offset order and excluding the normalization kernels of 7%, 4%, 2%, and 6%. In both data sets, as kernel offset increases, there is an increased delay before appreciable displacement is observed.

In the lateral displacement data comparison (below in Fig. 10), significant phenomena are observed. Both data sets present, independent of kernel offset, seemingly instantaneous and monotonically increasing displacement following ARFI excitation transmission. Additionally, with increasing kernel offset, peak displacement values begin increasing, reach an absolute peak (the normalization kernel), and then start decreasing in both data sets. Although TTPD values are not coincident, particularly at the two most distal kernels, they increase with increasing kernel offset in both the FEM and experimental cases. Percent differences, by increasing kernel offset, for TTPD values are 3%, 8%, 11%, 15%, and 13%, while percent differences in peak amplitude are 28%, 2%, 1%, 3%, and 10%. On average for the five kernels, non-normalized, peak axial displacements are 13 times larger than non-normalized, peak lateral displacements for both data sets.

IV. DISCUSSION

It was possible to accurately track embedded microspheres at depths nearing 5 mm, as evidenced by CCoef values typically above 0.85. Tracking reliability generally decreased (as indicated by reduced CCoef values) with increased FOV depth, as a result of increased optical scattering, and with increased frame rates, due to decreased SNR of the camera's sensor chip. The two greatest sources of experimental error were likely a result of inhomogeneity and registration issues. Inhomogeneities in the phantom material and/or beam intensity field accounted, in part, for discrepancies in observed dynamics between experiments or experimental trials. Temporally variant registration of a kernel and its corresponding microsphere or misregistration of the transducer/microscope foci resulted in inaccuracies in the estimation and spatial registration of induced displacements.

Material inhomogeneities in the phantom can arise during the production process or the experimental protocol itself. Given the inherent inhomogeneity of egg white solution, it is unlikely that a phantom with it as a principle base can be perfectly homogeneous. Such obviously inhomogeneous regions were specifically precluded from being experimental ROIs. In future experiments, the egg white solution will be filtered prior to its use in an attempt to further mitigate this effect. Material inhomogeneity can also result if egg white protein denatures due to heating generated from the microscope illumination source. In fact, it was for this very property that the basic phantom recipe was first employed by Takegami *et al.*[39] Although limiting illumination source use reduced this effect (with temperature increases $\leq 0.4\ ^{\circ}\mathrm{C}$ measured from illumination source heating alone), occasionally an ROI would become noticeably discolored as result of it, which required the designation of a new, neighboring ROI. Given the relatively short duration of the ARFI excitation pulses (≤ 0.4 ms) and the relatively long pulse repetition frequency (≥ 1 min), phantom heating due to absorption of acoustic radiation is thought to be insignificant. Using similar excitation pulses in a comparable medium, Palmeri *et al.*[45] estimated internal heating to be only a few tenths of a degree. Beam intensity "inhomogeneity" (i.e., deviation from the expected, axisymmetric transmission field of a circular piston) was either due to irregularities in the transmission path (as just discussed) or boundary condition artifacts. Reflections (of the longitudinal acoustic wave) at various interfaces (e.g., water/phantom or phantom/acrylic holder) could have resulted in significant aberration or attenuation of the acoustic intensity distribution at the ROI. Such focusing errors could have a substantial impact on the ARFI-induced response, as suggested by Eq. (3).

Registration issues accounted, in part, for inaccuracies in displacement characterization. As tracking CCoef values decreased, the registration between a microsphere and its corresponding kernel would worsen relative to the pre-excitation reference frame. Generally, this registration error manifested itself as tracking jitter while occasionally more significant displacement estimation inaccuracies resulted (e.g., the artifactual negative displacement in Fig. 5). Although the lowest CCoef values were achieved during transmission of the ARFI excitation (likely due to either in-plane motion blurring or significant out-of-plane motion of the microsphere), CCoef values were also noticeably affected by the camera sensor chip's SNR, which depended on frame-rate, sensitivity, and gain settings.

A second source of displacement characterization error originates from misregistration between the transducer and microscope foci. Given that the tip diameter of the needle hydrophone is on the same order as the push transducer's FWHM beamwidth, perfect lateral/elevational alignment is not possible. Additionally, elevational alignment of the hydrophone center with the microscope focal plane is not adequately achieved by merely ensuring that the needle tip is visibly in focus. This alignment technique can only ensure, assuming perfect elevational alignment between the hydrophone and push transducer foci, that the transducer's elevation focus is within a hydrophone tip radius (200 μm) of the microscope's imaging plane. Clear evidence of lateral misregistration is present in two of the data sets: the cyan kernel in Fig. 4 and the blue kernel in Fig. 7(a). In the first case, the cyan kernel peaked latest and achieved the greatest lateral displacement, despite not having the greatest reported lateral offset, suggesting that it was actually furthest from the excitation axis. In the second case, the sign of the blue kernel's lateral displacement is opposite the sign of its reported lateral position, implying that it moved *toward* the excitation center, which is physically implausible. Both inconsistencies are likely due to imperfect lateral registration between the FOV's and push transducer's lateral centers. Variation in axial position could have also introduced error in direct comparisons of different kernels. Yet, with the greatest range of axial positions in any data set only 340 μm, it is unlikely

that such error was significant given the much larger axial extent of the excitation beam.

Discrepancies between FEM modeling and experimental results are largely due to the misregistration and focusing issues detailed previously. In all FEM modeling results, the FOV lateral/elevational center was assumed to be perfectly coincident with the ARFI excitation lateral/elevational center. As already explained, such an assumption is not fully accurate. Thus, an improved characterization of a ROI's precise three-dimensional location would likely improve simulation agreement. Additionally, with broader excitation volumes generally resulting in a prolonged recovery period when compared to their narrower counterparts, differences in the effective beamwidth of the ARFI excitation would account, in part, for the discrepancy in recovery profiles observed between FEM and experimental axial displacement data.[24] The extended recovery phase in the experimental data might also have been influenced by boundary conditions (e.g., the proximal boundary) not properly accounted for in the FEM model. Additionally, the viscoelastic nature of the phantom—as opposed to the purely elastic behavior assumed in the model—could have influenced dynamics, particularly for those observed in the recovery profile. Thus, inclusion of viscosity into the FEM model could also improve agreement in the future.

The off-axis tracking data offered perhaps the most interesting findings. In the lateral displacement data, the seemingly instant displacement observed in all off-axis, lateral displacement plots [Figs. 7(b)–7(d) and below in Fig. 10] is likely due to a Poisson effect, which requires instant, global redistribution of an incompressible medium when a local strain field is applied to it. An ARFI excitation causes mass within its beam volume to displace away from the push transducer, with preferential force applied (resulting in increased displacement) to mass located near the beam's focus. When this focal volume displaces away from the transducer, surrounding phantom material must redistribute due to its nearly incompressible nature. Material deep to the focus must displace away from the central beam axis to "make room" for the newly displaced volume while material shallow to the focus must displace toward the beam axis to "fill in" for this volume. Given that the phantom was not perfectly incompressible, this redistribution does not happen instantaneously but rather at a speed which is orders of magnitude faster than the shear wave velocity.

In the off-axis, axial displacement data [Figs. 6(b)–6(d)], the progressive dispersion and attenuation observed in PBSW reflection TTPD and peak displacement values, respectively, are due to the increasing shear wave propagation distance to more distal kernels. If the shear wave reflection is thought to have emanated from the excitation's image across the proximal boundary (so hypothetically near the microscope's objective), it becomes clear that reflected wave propagation distances will be least for kernels nearest the axis connecting the source/image foci. As kernels are positioned further from this axis, reflected shear wave propagation distances increase, which results in later TTPD values and greater peak attenuations at those times.

In the future, it is hoped that significant improvements can be made in design of the optical phantom. To increase the number of trackable microspheres within the FOV, a highly populated "monolayer" of microspheres can be generated in the phantom at a specific depth from the proximal boundary. This can be achieved by superficially applying a layer of microspheres on the phantom's original proximal boundary then casting a thin, matched layer atop the microspheres to form a new proximal boundary; this is a similar concept to the dual-stage casting technique employed by Andreev *et al.*[31] This multi-staged fabrication process can then be taken a step further by adding a third, minimally scattering (optically), mechanically matched layer atop the second layer in an attempt to increase the effective proximal boundary distance without drastically compromising microscope imaging quality. A gelatin layer without egg white or microspheres could be nearly transparent (which would increase light transmission and improve focusing ability) and formulated to match the stiffness of the other two layers, which would eliminate (or severely mitigate) the creation of a shear wave reflection at that boundary and would extend the PBSW reflection distance to the more distal boundary of this third layer.

It is also the authors' intention to construct mechanically inhomogeneous optical phantoms for the purpose of investigating shear wave dynamics at material interfaces. If these second and third layers were cast from a higher bloom strength gelatin (or if two different gelatin phases were created in the transverse plane, in alignment with the microscope axis), there would be an appreciable mechanical contrast at this interface. It would then be possible to investigate (optically) how a shear wave propagates along or couples into such a boundary. With regard to the experimental setup, improvements to the transducer micro-positioning system could be made to allow for closer (relative to the microscope's FOV) transducer placement; this would facilitate optical investigation of the dynamic response shallow to the excitation focus. Additionally, the push transducer, or an added opposing but coaxial "tracking" transducer, could be utilized for the purpose of synchronized and matched ultrasonically based tracking of the post-ARFI response, similar to the approach implemented previously by Bouchard *et al.*[36]

Despite the inhomogeneity and registration issues addressed herein, the presented optically based method is capable of accurately tracking, with improved spatial and temporal resolutions, the dynamic response of a tissue-mimicking phantom at depth. Although the phantom's proximal boundary introduces a clear artifact later in time, its influence is predictable [Fig. 8(a)] and does not appear to affect the final recovery of the phantom [Fig. 8(b)]. Post-excitation, axial displacement results from similar FEM models have previously been validated with experimental data obtained in a semi-infinite phantom environment.[16] Thus, corroboration of experimentally observed dynamical phenomena and displacement trace morphologies with FEM results suggest that the optical phantom setup, early in time, is able to facilitate similar ARFI-induced dynamics as those generated in a semi-infinite medium. Additionally, independent experimental validation of optical tracking is offered

through ultrasonically based shear wave velocimetry (Exp. 8), which yielded a velocity estimate that is not statistically differentiable from the one provided by the optically based technique.

V. CONCLUSION

Optical tracking of ARFI-induced dynamics in a tissue-mimicking phantom was successfully achieved at frame rates of up to 36 kHz and with sub-micron displacement resolution in the axial and lateral dimensions. These tracking data show good agreement with all basic trends and phenomena observed in matched FEM modeling results early in time (up to 4 ms). Excellent agreement is also observed between shear wave velocities derived from the optical technique and those yielded by an independent ultrasonically based method. Due to the closeness of the phantom's proximal boundary, an artifact-generating shear wave reflection was observed in all data sets later in time (around 5 ms in most cases shown). It is hoped that this artifact as well as the limited number of tracking kernels available within a FOV can be addressed in the future with an improved phantom design. Despite the restricted clinical applicability of this tracking technique, it could assist in gaining a greater understanding of complex radiation force dynamics in tissue-mimicking phantoms.

ACKNOWLEDGMENTS

This work was supported in part by a National Science Foundation Graduate Research Fellowship. The authors thank Gijs van Soest, Carl Herickhoff, and Michael Giacomelli for their insight and Ned Rouze for intensity measurement assistance.

[1]K. R. Nightingale, M. L. Palmeri, R. W. Nightingale, and G. E. Trahey, "On the feasibility of remote palpation using acoustic radiation force," J. Acoust. Soc. Am. **110**, 625–634 (2001).
[2]D. Melodelima, J. Bamber, F. Duck, J. Shipley, and L. Xu, "Elastography for breast cancer diagnosis using radiation force: System development and performance evaluation," Ultrasound Med. Biol. **32**, 387–396 (2006).
[3]Y. A. Ilinskii, G. D. Meegan, E. A. Zabolotskaya, and S. Y. Emelianov, "Gas bubble and solid sphere motion in elastic media in response to acoustic radiation force," J. Acoust. Soc. Am. **117**, 2338–2346 (2005).
[4]R. H. Behler, T. C. Nichols, H. Zhu, E. P. Merricks, and C. M. Gallippi, "ARFI imaging for noninvasive material characterization of atherosclerosis Part II: Toward *in vivo* characterization," Ultrasound Med. Biol. **35**, 278–295 (2009).
[5]J. Bercoff, M. Tanter, and M. Fink, "Supersonic shear imaging: A new technique for soft tissue elasticity mapping," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 396–409 (2004).
[6]A. Sarvazyan, O. Rudenko, S. Swanson, J. Fowlkes, and S. Emelianov, "Shear wave elasticity imaging: A new ultrasonic technology of medical diagnostics," Ultrasound Med. Biol. **24**, 1419–1435 (1998).
[7]S. A. McAleavey, M. Menon, and J. Orszulak, "Shear-modulus estimation by application of spatially-modulated impulsive acoustic radiation force," Ultrason. Imaging **2**, 87–104 (2007).
[8]E. E. Konofagou and K. Hynynen, "Localized harmonic motion imaging: Theory, simulations and experiments," Ultrasound Med. Biol. **29**, 1405–1413 (2003).
[9]M. Fatemi and J. F. Greenleaf, "Vibro-acoustography: An imaging modality based on ultrasound-stimulated acoustic emission," Proc. Natl. Acad. Sci. U.S.A. **96**, 6603–6608 (1999).
[10]F. Viola and W. F. Walker, "Radiation force imaging of viscoelastic properties with reduced artifacts," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 736–742 (2003).
[11]J. Ophir, I. Céspedes, H. Ponnekanti, Y. Yazdi, and X. Li, "Elastography: A quantitative method for imaging the elasticity of biological tissues," Ultrason. Imaging **13**, 111–134 (1991).
[12]K. Hoyt, T. Kneezel, B. Castaneda, and K. J. Parker, "Quantitative sonoelastography for the *in vivo* assessment of skeletal muscle viscoelasticity," Phys. Med. Biol. **53**, 4063–4080 (2008).
[13]S. J. Hsu, R. R. Bouchard, D. M. Dumont, P. D. Wolf, and G. E. Trahey, "*In vivo* assessment of myocardial stiffness with acoustic radiation force impulse imaging," Ultrasound Med. Biol. **33**, 1706–1719 (2007).
[14]J. L. Gennisson, S. Catheline, S. Chaffaï, and M. Fink, "Transient elastography in anisotropic medium: Application to the measurement of slow and fast shear wave speeds in muscles," J. Acoust. Soc. Am. **114**, 536–541 (2003).
[15]H. Kanai, "Propagation of spontaneously actuated pulsive vibration in human heart wall in *in vivo* viscoelasticity estimation," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 1931–1942 (2005).
[16]M. L. Palmeri, A. C. Sharma, R. R. Bouchard, R. W. Nightingale, and K. R. Nightingale, "A finite-element method model of soft tissue response to impulsive acoustic radiation force," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 1699–1712 (2005).
[17]P. J. Westervelt, "The theory of steady forces caused by sound waves," J. Acoust. Soc. Am. **23**, 312–315 (1951).
[18]M. E. Lyons and K. J. Parker, "Absorption and attenuation in soft tissues. II. Experimental results," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 511–521 (1988).
[19]W. Nyborg, "Acoustic streaming," in *Physical Acoustics*, edited by W. Mason (Academic, New York, 1965), Vol. **IIB**, Chap. 11, pp. 265–331.
[20]G. R. Torr, "The acoustic radiation force," Am. J. Phys. **52**, 402–408 (1984).
[21]W. Lai, D. Rubin, and E. Krempl, *Introduction to Continuum Mechanics* (Butterworth-Heinemann, Woburn, MA, 1999).
[22]J. Bishop, G. Poole, and D. Plewes, "Magnetic resonance imaging of shear wave propagation in excised tissue," J. Magn. Reson. Imaging **8**, 1257–1265 (1998).
[23]M. Fatemi and J. F. Greenleaf, *Topics in Applied Physics* (Springer Berlin, Heidelberg, 2002), Vol. **84**, pp. 257–276.
[24]M. L. Palmeri, S. A. McAleavey, K. L. Fong, G. E. Trahey, and K. R. Nightingale, "Dynamic mechanical response of elastic spherical inclusions to impulse acoustic radiation force excitation," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 2065–2079 (2006).
[25]G. F. Pinton, J. J. Dahl, and G. E. Trahey, "Rapid tracking of small displacements with ultrasound," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 1103–1117 (2006).
[26]M. A. Lubinski, S. Y. Emelianov, K. R. Raghavan, A. E. Yagle, A. R. Skovoroda, and M. O'Donnell, "Lateral displacement estimation using tissue incompressibility," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 247–256 (1996).
[27]R. R. Bouchard, J. J. Dahl, S. J. Hsu, M. L. Palmeri, and G. E. Trahey, "Image quality, tissue heating, and frame rate trade-offs in acoustic radiation force impulse imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **56**, 63–76 (2009).
[28]C. Maleke, M. Pernot, and E. E. Konofagou, "Single-element focused ultrasound transducer method for harmonic motion imaging," Ultrason. Imaging **28**, 144–158 (2006).
[29]M. L. Palmeri, S. A. McAleavey, G. E. Trahey, and K. R. Nightingale, "Ultrasonic tracking of acoustic radiation force-induced displacements in homogeneous media," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 1300–1313 (2006).
[30]J. Schmitt, "OCT elastography: Imaging microscopic deformation and strain of tissue," Opt. Express **3**, 199–211 (1998).
[31]V. G. Andreev, V. N. Dmitriev, Y. A. Pischal'nikov, O. V. Rudenko, O. A. Sapozhnikov, and A. P. Sarvazyan, "Observation of shear waves excited by focused ultrasound in a rubber-like media," Acoust. Phys. **43**, 123–128 (1996).
[32]E. Bossy, A. R. Funke, K. Daoudi, A.-C. Boccara, M. Tanter, and M. Fink, "Transient optoelastography in optically diffusive media," Appl. Phys. Lett. **90**, 174111 (2007).
[33]P. A. Dayton, K. E. Morgan, A. L. Klibanov, G. Brandenburger, K. R. Nightingale, and K. W. Ferrara, "A preliminary evaluation of the effects of primary and secondary radiation forces on acoustic contrast agents," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 1264–1277 (1997).
[34]P. Palanchon, P. Tortoli, A. Bouakaz, M. Versluis, and N. de Jong, "Optical observation of acoustical radiation force effects on individual air bubbles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 104–110 (2005).
[35]P. A. Dayton, J. S. Allen, and K. W. Ferrara, "The magnitude of radiation force on ultrasound contrast agents," J. Acoust. Soc. Am. **112**, 2183–2192

(2002).
[36]R. R. Bouchard, G. van Soest, G. E. Trahey, and A. F. W. van der Steen, "Optical tracking of superficial dynamics from an acoustic radiation force-induced excitation," Ultrason. Imaging **31**, 17–30 (2009).
[37]M. C. W. van Rossum and T. M. Nieuwenhuizen, "Multiple scattering of classical waves: Microscopy, mesoscopy, and diffusion," Rev. Mod. Phys. **71**, 313–371 (1999).
[38]P. Sheng, *Introduction to Wave Scattering, Localization, and Mesoscopic Phenomena* (Academic, New York, 1995).
[39]K. Takegami, Y. Kaneko, T. Watanabe, T. Maruyama, Y. Matsumoto, and H. Nagawa, "Polyacrylamide gel containing egg white as new model for irradiation experiments using focused ultrasound," Ultrasound Med. Biol. **30**, 1419–1422 (2004).
[40]T. J. Hall, M. Bilgen, M. F. Insana, and T. A. Krouskop, "Phantom materials for elastography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 1355–1365 (1997).
[41]F. W. Kremkau, R. W. Barnes, and P. McGraw, "Ultrasonic attenuation and propagation speed in normal human brain," J. Acoust. Soc. Am. **70**, 29–38 (1981).
[42]M. L. Palmeri, M. H. Wang, J. J. Dahl, K. D. Frinkley, and K. R. Nightingale, "Quantifying hepatic shear modulus *in vivo* using acoustic radiation force," Ultrasound Med. Biol. **34**, 546–558 (2008).
[43]J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 262–267 (1992).
[44]J. A. Jensen, "Field: A program for simulating ultrasound systems," in Proceedings of the 10th Nordic-Baltic Conference on Biomedical Imaging (1996), Vol. **34**, pp. 351–353.
[45]M. Palmeri and K. Nightingale, "On the thermal effects associated with radiation force imaging of soft tissue," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 551–565 (2004).

Shock-induced bubble jetting into a viscous fluid with application to tissue injury in shock-wave lithotripsy

J. B. Freund[a)]
Mechanical Science and Engineering and Aerospace Engineering, University of Illinois at Urbana-Champaign, 1206 West Green Street, MC-244, Urbana, Illinois 61801

R. K. Shukla
Center for Simulation of Advanced Rockets, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

A. P. Evan
Department of Anatomy and Cell Biology and Medicine, Indiana University School of Medicine, Indianapolis, Indiana 46202

(Received 22 April 2009; revised 18 August 2009; accepted 19 August 2009)

Shock waves in liquids are known to cause spherical gas bubbles to rapidly collapse and form strong re-entrant jets in the direction of the propagating shock. The interaction of these jets with an adjacent viscous liquid is investigated using finite-volume simulation methods. This configuration serves as a model for tissue injury during shock-wave lithotripsy, a medical procedure to remove kidney stones. In this case, the viscous fluid provides a crude model for the tissue. It is found that for viscosities comparable to what might be expected in tissue, the jet that forms upon collapse of a small bubble fails to penetrate deeply into the viscous fluid "tissue." A simple model reproduces the penetration distance versus viscosity observed in the simulations and leads to a phenomenological model for the spreading of injury with multiple shocks. For a reasonable selection of a single efficiency parameter, this model is able to reproduce *in vivo* observations of an apparent 1000-shock threshold before wide-spread tissue injury occurs in targeted kidneys and the approximate extent of this injury after a typical clinical dose of 2000 shock waves. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3224830]

PACS number(s): 43.80.Jz, 43.35.Wa, 43.40.Ng [ROC] Pages: 2746–2756

I. INTRODUCTION

We consider a small gas-filled bubble being compressed rapidly by a shock wave (see Fig. 1) and its subsequent jetting toward a viscous material. This configuration is motivated by medical procedures such as shock-wave lithotripsy, during which shock waves are directed toward kidney stones in the hope of fracturing them into "passable" pieces. At clinical shock-wave doses, there appears to be significant collateral injury to the kidney,[1,2] which is implicated in certain short- and long-term complications.[3] The action of cavitation bubbles is implicated in this injury.[4,5]

Bubble expansion, caused by the negative-pressure phase of the lithotripter wave,[6] has been suggested as a potential mechanism of the injury,[7] but the bubble collapse is also potentially damaging. It is known that a bubble can collapse asymmetrically leading to the formation of a so-called re-entrant jet,[8,9] which starts from where the shock first encounters the bubble and is able to penetrate the bubble's far side with sufficient velocity to damage nearby material. This is one of the mechanisms thought to cause cavitational damage in engineered systems in cases where the flow's dynamic pressure causes the cavitation and subsequent collapse.[8] The shock sensitivity of explosives also appears to depend on this jetting mechanism. In this case, the formation of local hot spots in the material by the dissipation associated with this jetting seems to increase the overall explosive sensitivity of energetic materials to shock-like mechanical impacts.[10,11]

In tissues, this jetting has been hypothesized to be the mechanism of mechanical injury during lithotripsy (e.g., see the recent discussion of Klaseboer *et al.*[12]), and it is potentially the mechanism by which bubbles subjected to bursts of high-intensity focused ultrasound (HIFU) can erode tissue (e.g., Ref. 13). HIFU is also well known to cause thermal injury to tissue, but our concern is with mechanical effects at energy deposition rates that preclude significant heating. Thermal injury is not expected in lithotripsy.[14]

Simulations of collapsing bubbles typically neglect viscosity,[12,15–21] which is indeed justified based on the Reynolds numbers of the jets expected under typical conditions,[20] though for very small bubbles viscous effects have been identified for non-shock-induced (so-called Rayleigh) collapse near a wall.[22] The re-entrant jets for lithotripter shocks appear to have speeds of around 1000 m/s,[12] so for a 1 mm diameter bubble in water the jet Reynolds number is about 10^6. Even if we assume that the re-entrant jet diameter is only 1% of the bubble diameter, this Reynolds number is still 10^4. However, the significantly smaller bubbles that might form in microvessels in the kidney (say, 20 μm diameter) and the significantly higher viscosities of

a)Author to whom correspondence should be addressed. Electronic mail: jbfreund@illinois.edu

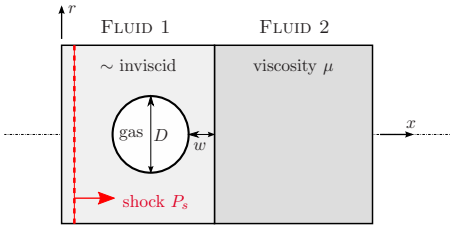


FIG. 1. (Color online) Configuration schematic (see text).

tissue (at least hundreds of times that of water) can lead to re-entrant jets with Reynolds numbers of around unity. This suggests that tissue viscosity might play a significant role in suppressing the jetting and any injury it might cause. Recent experiments involving laser-induced bubble growth and collapse in viscous fluids suggest that higher viscosity fluids both suppress the strength of the jetting and slow the time scale of the collapse.[23]

Viscosity has also recently been proposed to be important for the confinement of bubble expansion when subjected to model lithotripter shock profiles.[24] Assuming spherical symmetry, we recently generalized the well-known Rayleigh–Plesset bubble dynamics model to account for confinement by an elastic membrane and an extensive Voigt visco-elastic material.[24] Results suggest that even the highest estimates of tissue elasticity fail to suppress bubble growth significantly, but because of the small scales and nature of the expansion, even moderate estimates of tissue viscosity were able to play a substantive role is suppressing bubble expansion.

Though kidney injury directly motivates this study, it should also be clear that connection of the present study to tissue and its injury is loose since we neglect its elastic character altogether and study the collapsing bubble's interaction with a viscous fluid. Our Newtonian viscosity can, of course, provide only an approximate model for tissue viscosity under small deformations and only the crudest possible model for the dissipation associated with the mechanical disruption of tissue. That understood, this type of interaction does not appear to have been studied and a linear viscosity model is clearly a reasonable start for investigating phenomenology. Extending the type of simulations employed herein to tissue-like visco-elasto-plastic models would be a non-trivial task. Furthermore, any such attempts to refine the tissue model would remain only phenomenological because of the lack of detailed constitutive models for the mechanics of tissue injury.

Section II introduces the physical model for studying jetting penetration of the viscous "tissue." The numerical scheme is summarized in Sec. III, and simulation results along with a simple model reproducing the jet penetration depth are presented in Sec. IV. In Sec. V, this simple model is incorporated into a phenomenological model for the spreading of tissue injury with the thousands of shocks of a typical treatment. There it is shown to be successful at reproducing some *in vivo* observations of kidney injury during lithotripsy.

II. THE MODEL CONFIGURATION

We are interested primarily in the effect of a pre-existing bubble as it collapses on an adjacent viscous fluid, particularly the penetration depth of the jet, which motivates the configuration shown in Fig. 1. The shock propagates perpendicular to the viscous fluid. The only length scales are the bubble diameter D and its distance w from fluid 2 which has viscosity μ. Both the viscous and nearly inviscid (μ_1) liquids have a density ρ that is 1000 times that of the gas in the bubble. This density ratio is so large that the density of the gas is expected to have negligible influence on the subsequent jetting dynamics. Simulations confirm that doubling it does not change any of the results we discuss. There are thus only two parameters we consider: $s=w/D$ and $\mathrm{Re}=\sqrt{P_s\rho}D/\mu$, the second of which is the Reynolds number based on shock pressure jump P_s and density ρ, which is commonly used in studies of shock-induced pore collapse (e.g., Ref. 25). For the nearly inviscid fluid, we take $\mathrm{Re}_1=\sqrt{P_s\rho}D/\mu_1=4000$, so we do not expect significant viscous effects in the collapse itself. Five Reynolds numbers are simulated (Re=0.4, 4, 13.3, 40, and 400), which we anticipate should show highly dissipative to relatively inviscid behavior. We also consider $s=0$, 0.25, and 0.5, and shall see in Sec. IV that the bubble collapse and peak jet velocity is insensitive to s for this range.

The non-dimensional parameters s and Re guide our investigation of the relative effects of bubble proximity and viscosity, but our simulations are also motivated by the specific bubble-in-tissue application. The parameters considered correspond approximately to a 20 μm diameter air bubble in water at atmospheric pressure being compressed by a $P_s=40$ MPa shock, as might be delivered by a typical lithotripter.[26] The liquid densities are both 1000 kg/m^3. The lowest tissue viscosity we consider is $\mu=0.01$ Pa s, which corresponds roughly to tissue viscosity deduced via the small fast deformations of dissipating ultrasonic shear waves.[27–29] Our results (Sec. IV) show that indeed there is little viscous dissipation in this case. The highest viscosity we consider is $\mu=10$ Pa s, which corresponds to more standard, high amplitude but slower rate, deformations.[30] This range of viscosities is discussed in more detail elsewhere.[24] Because of the current activity in this specific area, we choose to present mostly dimensional results for this particular system.

It is known that the shock might couple with on-going oscillations of the bubble, affecting both collapse time and apparent jet strength,[31] which has been studied in some detail via boundary integral simulation methods.[12] However, to simplify our investigation we consider the bubble to be of fixed volume before the interaction with the shock. We anticipate that the jet formation will occur almost independently of its subsequent interaction with the viscous half-space, which is indeed confirmed in Sec. IV. So, assuming the oscillations are relatively weak, the current jet penetration results should apply to a transient or oscillating bubble so long as the jet strength is properly accounted for.

The equations governing the system are

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \tag{1}$$

$$\frac{\partial \rho \mathbf{u}}{\partial t} + \nabla \cdot (\rho \mathbf{u}\mathbf{u}) + \nabla p = \nabla \cdot \tau, \tag{2}$$

$$\frac{\partial E}{\partial t} + \nabla \cdot [\mathbf{u}(E+p)] = \nabla \cdot (\tau \cdot \mathbf{u}), \tag{3}$$

where ρ is the density, p is the pressure, $\mathbf{u}$ is the velocity, E is the total energy (kinetic plus internal), and τ are the Newtonian-fluid viscous stresses. These equations are written compactly here in three-dimensions, but the simulation model is constrained to be axisymmetric. The liquid thermodynamics are modeled via a stiffened equation of state,

$$p = \rho(\gamma - 1)e - \gamma p_\infty, \tag{4}$$

where e is the mass-specific internal energy. The model parameters in Eq. (4) are selected to be $p_\infty = 3 \times 10^8$ Pa and $\gamma = 7$, which provides an approximate model for water.[32] The bubble contents are modeled by the ideal gas equation of state: Eq. (4) with $p_\infty = 0$ and $\gamma = 1.4$.

Here we consider only the collapse and jetting process for a pre-existing spherical bubble. In lithotripsy, these are expected to be due to gas being pulled out of solution by the negative portion of the lithotripter wave, which follows the shock.[4,6] The numbers of such bubbles are also expected to increase with subsequent shocks and the intensity of shock-induced collapse has motivated a close examination of their potential role in stone breaking.[19] The jetting in this case also seems to be somewhat stronger than in corresponding Rayleigh collapse.[20] While dissolution and diffusion are expected to be important for the formation of the gas bubbles, the collapse and jetting are expected to occur too quickly for any significant phase change or thermal transfer between the bubble and the surrounding fluid to affect the subsequent jet formation. Inclusion of such factors is important for calculating peak temperatures for asymmetric collapses,[33] and presumably for behavior of the collapsed bubble upon re-expansion, but the jet formation is expected to be governed by mass and momentum conservation considerations that are independent of the microscopic details of the final collapsed condition. Simulation models that do not attempt to resolve the heat transfer, diffusion, and high-temperature effects in the equation of state behavior have indeed shown collapse dynamics that seem to match corresponding experimental data (e.g., Ref. 16). We follow this approach and neglect these factors, though our simulation does include liquid compressibility and viscosity, which are also often omitted from simulations. We also neglect surface tension. Weber numbers $\mathrm{We} = \rho U^2 D/\sigma$ based on bubble diameter, peak calculated jet velocities (see Sec. IV), and the surface tension of blood [$\sigma = 0.06$ N/m (Ref. 34)] are around 5×10^5 and would be roughly the same for water. Thus, inertia is clearly expected to dominate surface tension for jet formation, justifying its neglect here.

For the small bubbles of interest, we shall see that the entire collapse and jetting penetration of the viscous fluid takes place in $\lesssim 0.25$ μs, with conclusions about suppressed penetration available after around 0.05 μs. For these times, a lithotripter shock can be assumed to have a sharp rise and then constant pressure. In water, where the pressure profiles are typically measured, the shock-wave pressure has an unresolvably fast rise time [theoretically around 0.15 ns (Ref. 35)] before it drops approximately linearly over about 1 μs.[26] So by the time there would any significant decrease in the wave pressure, the bubble in our simulations will have collapsed to such a small size that its dynamics will not affect the subsequent jetting.

III. NUMERICAL SOLVER

The basic simulation approach is similar to that of Johnsen and Colonius,[18] who also considered bubble collapse by lithotripter shocks, though our objectives and the details of our algorithm are different. Our finite-volume solver for Eqs. (1)–(4) uses a TVD reconstruction with a minmod limiter[36] and a HLLC (Ref. 37) approximate Riemann solver. These are standard techniques for single-phase gas dynamics calculations involving shocks. To track the three fluids in our system (the gas, fluid 1, and fluid 2), we also transport two phase variables, which are used to demark the different regions that the different fluids occupy: $\phi_g = 1$ in the bubble and is 0 elsewhere, and $\phi_s = 1$ in fluid 2 and is zero elsewhere. A wide class of level-set or phase-field schemes models the interfaces, which in reality are molecularly thin, with a mesh-resolved though narrow continuous variation of ϕ between its extreme values on either side. With such a "smeared" interface model, the transport of the different ϕ's is governed by

$$\frac{\partial \phi_g}{\partial t} + \mathbf{u} \cdot \nabla \phi_g = 0, \tag{5}$$

$$\frac{\partial \phi_s}{\partial t} + \mathbf{u} \cdot \nabla \phi_s = 0. \tag{6}$$

Numerical diffusion in general will further smear these interfaces in time, which greatly degrades the quality of long-time solutions. However, we have designed special terms based on initialization of the phase field using a tanh profile diffused over a few grid cells and keeping this profile fixed as it advects during the simulation.[38,39] When coupled into the overall numerical scheme, this preserves the sharpness of the ϕ representation of the material boundaries. These and all the details of the inviscid portion of the scheme are presented in full elsewhere.[40] Viscous terms appearing in Eqs. (2) and (3) are discretized using a standard second-order finite-volume scheme in a way that keeps the overall method conservative. A four-stage third-order accurate semi-implicit Runge–Kutta method[41] is used to treat viscous terms implicitly and effectively avoids the strong stability restriction encountered by explicit time integration methods for the higher viscosities. The resulting coupled system of linear equations for momentum in the x and r directions is solved using the BICGSTAB (Ref. 42) algorithm.

Solutions with finite-volume solvers are not potentially as fast as with boundary-element methods,[16] but the inclusion of viscosity seems to preclude boundary-only discretizations since Green's functions are only available in the inviscid and strictly viscous limits. Boundary integrals would also be inconsistent with our current simulations in which we wish to track the fluid jet even after the bubble has collapsed to very small (negligible in our model) size. Perhaps not essential, but potentially important, a finite-volume solver

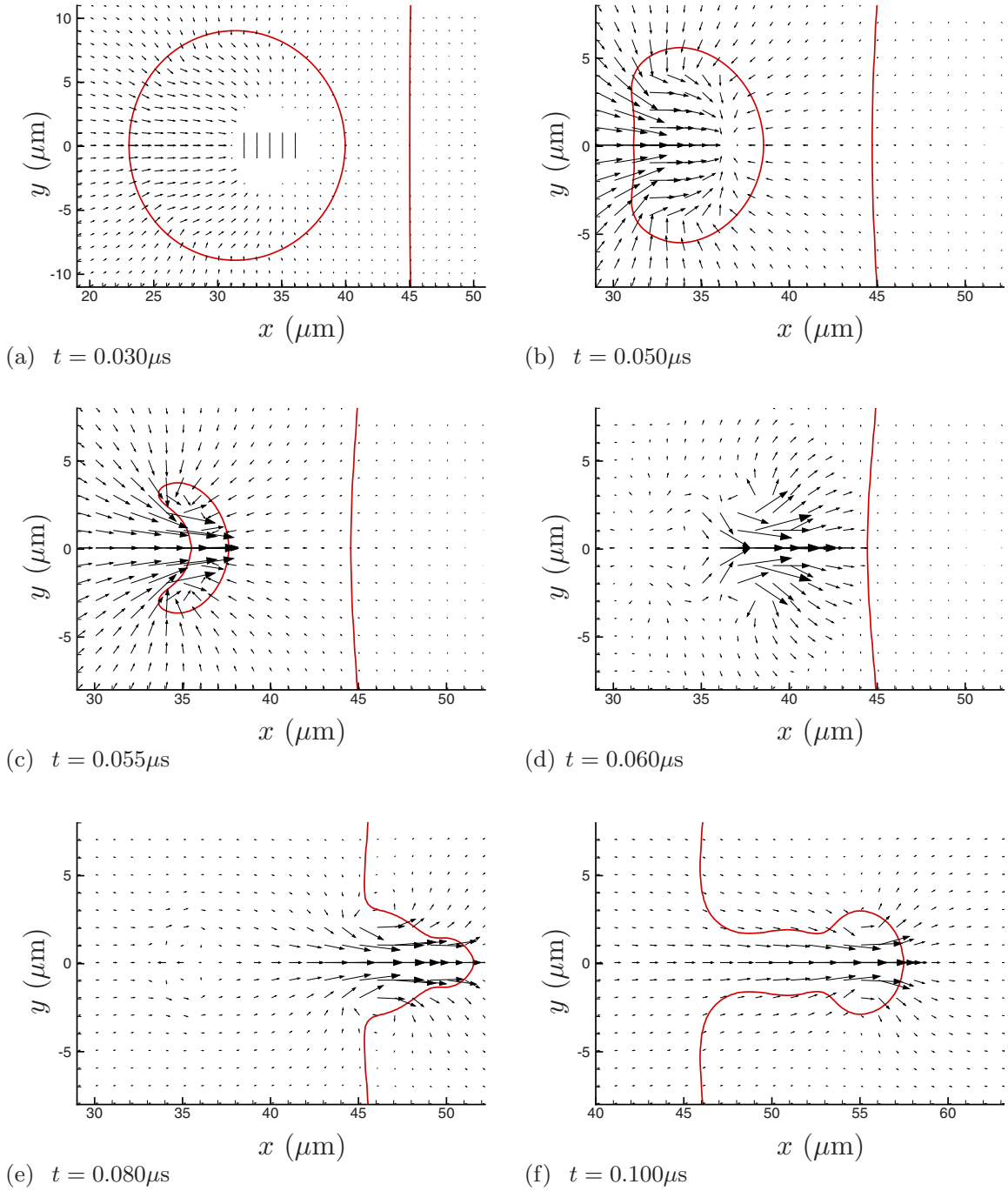


FIG. 2. (Color online) Velocity vectors at times during the collapse and jetting for the μ=0.1 Pa s case with $w=D/4=5$ μm. Note that the scale is different in (a) to show the full bubble at the beginning of its collapse. The lines show the $\rho=0.5\rho_{\text{liquid}}$ and $\phi_f=0.5$ isopleths. The vector associated with only every eighth mesh point in each coordinate direction is plotted. For reference, in (a), the shock front is just leaving the region shown.

also allows explicit two-way coupling of the shock propagation and the bubble response. The simulations presented herein required 1–3 processor-days to complete, which is still very fast given the availability of parallel systems. The longest of times were required for the most viscous cases because of longer convergence times for the implicit time advancement.

The axisymmetric computational domain extends 100 μm in x and out to 50 μm in r and was discretized by 800×400 mesh points in these two directions, respectively. For all the penetration distances calculated, less than a 10% change was observed for a 400×200 mesh calculation. The fluid 1/fluid 2 boundary was at $x=45$ μm and the shock was initialized using the shock-jump conditions for Eq. (4) at $x=10$ μm. Simulations were run with time step Δt adjusted to fix the CFL number: $\Delta t(c+|u|)_{\max}/\Delta x=0.3$, where Δx is the mesh spacing.

IV. RESULTS

Figures 2–4 visualize the collapse and re-entrant jetting. For the higher viscosity cases (Fig. 3 and especially Fig. 4), the penetration of the jet into fluid 2 appears to be suppressed. But it is also clear from frame (c) of all three figures

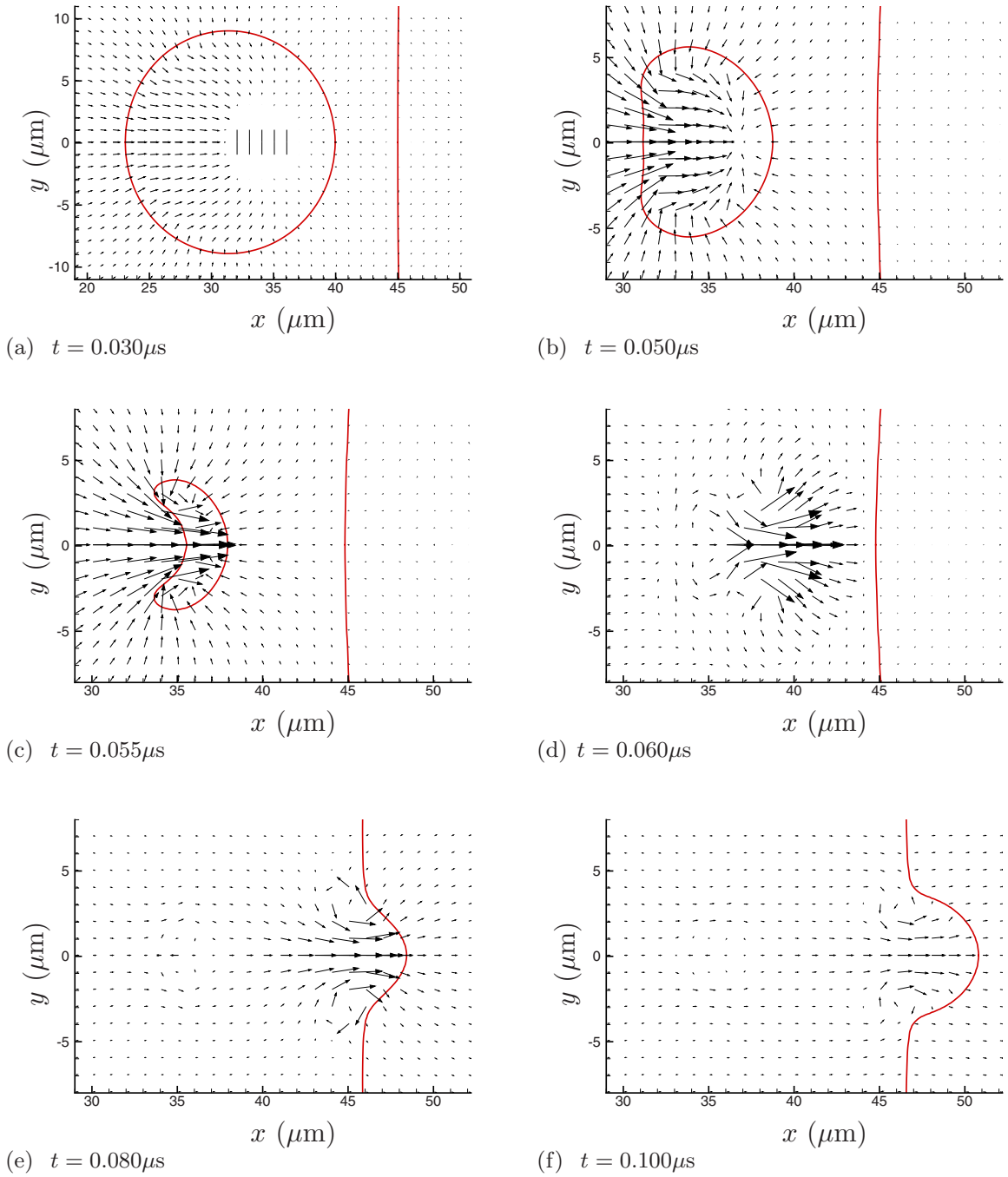


FIG. 3. (Color online) Same as Fig. 2 for the μ=1 Pa s case.

that the initial asymmetric collapse and jet formation is insensitive to the viscosity of fluid 2. Quantitative comparisons of velocities confirm this. Peak x-direction velocities through the entire collapse and jet formation are nearly the same for all cases, as can be seen in Table I, though the viscous resistance of fluid 2 does increase jet velocities a bit. (The peak jet velocity of 1.2 km/s is similar to that found for similar shocks by others.[19,43]) The influence of the viscous fluid 2 is strongest when the bubble is directly in contact with fluid 2 (the w=0 case) up through μ=1 Pa s, though this trend reverses for the highest viscosity. Except perhaps for the w =0 case, we can conclude that the collapse and initial jet formation are insensitive to the viscosity of fluid 2. A more significant increase in jetting velocity is seen for the case of a bubble adjacent a solid wall,[19,43] which is presumably mostly due to the reflection of the shock. Our results, in which there is no acoustic impedance mismatch, suggest that there might also be a relatively small hydrodynamic influence from the wall. In all cases, the bubble collapses as it should to a size that is too small to be resolved by the numerical scheme, but because of this small size and mass it will also not affect the subsequent jetting dynamics.

Our principal interest is the distance $d(t)$ to which the jet penetrates the viscous material, because this is presumably related to disruption of tissue and the spreading of injury. To account for the small uniform advection due to the post-shock velocity, $d(t)$ is taken to be the x distance between the

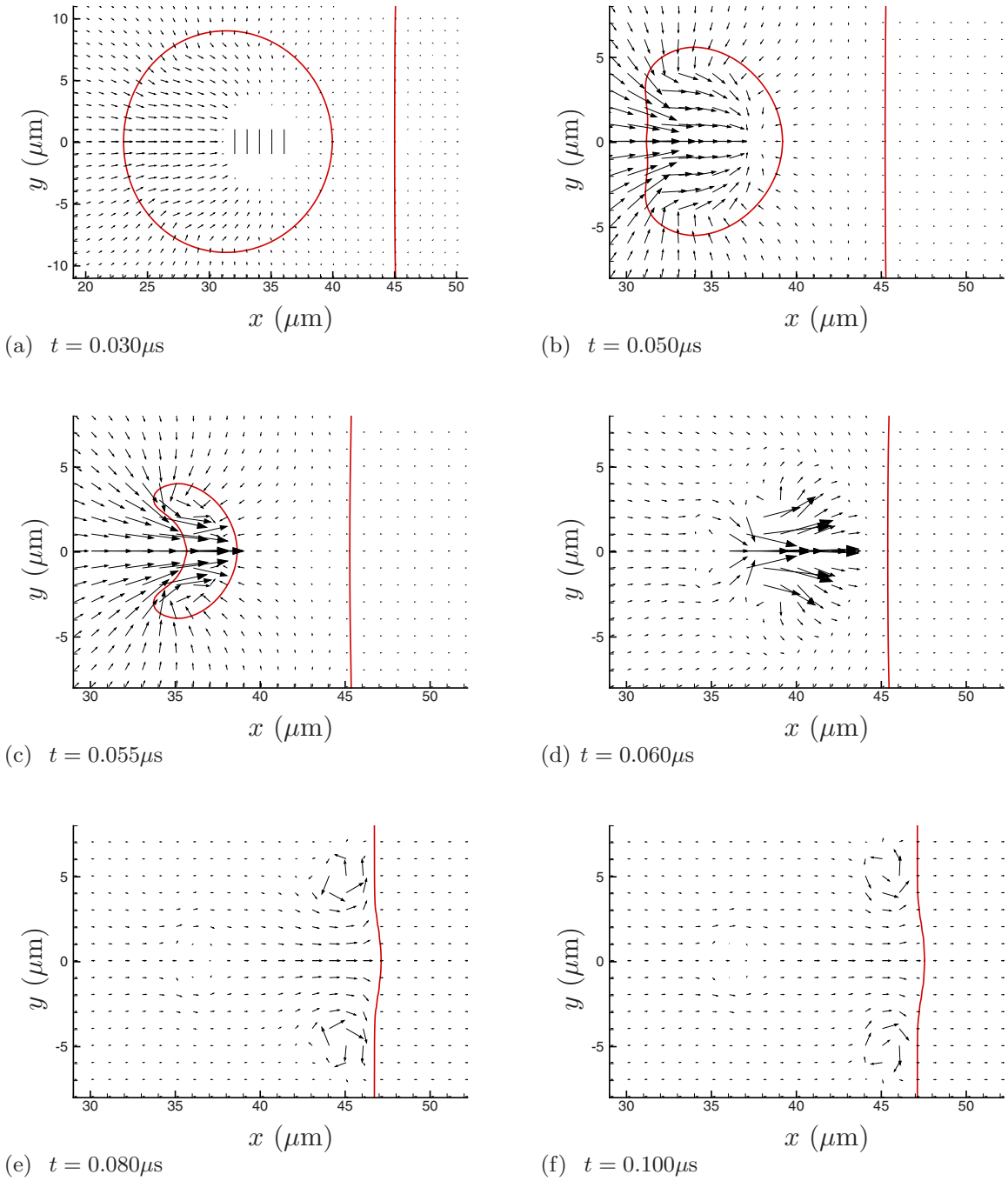


FIG. 4. (Color online) Same as Fig. 2 for the μ=10 Pa s case.

fluid 1/fluid 2 interface at the r=50 μm, where the post-shock x velocity is steady and uniform, and the interface location at r=0. This penetration depth is plotted in Fig. 5(a). In all cases, our metric shows that the interface-bubble separation first decreases, an effect which is particularly pronounced for the w=0 cases. It is pulled this way by the initial collapse, which is relatively symmetric at first, and because the bubble shields the interface at r=0 from the shock's acceleration, which has been noted previously in the context of interaction with kidney stones.[19] This can be seen in the (b) frames of Figs. 2–4. After the initial attraction of the interface toward the collapsing bubble, Fig. 5(a) shows that increasing viscosity substantially slows the jet and suppresses its penetration into fluid 2. The minor changes in peak jet velocity seen in Table I do not lead to significant differences in penetration depth, nor do the different distances between the bubble and the wall. The penetration increases linearly in time for the lowest viscosity cases, which shows that it is

TABLE I. Peak jet velocities in km/s for all cases.

μ (Pa s)	w=0	w=D/4	w=D/2
0.01	1.27	1.28	1.29
0.10	1.37	1.30	1.29
1.30	1.46	1.31	1.30
1.00	1.60	1.30	1.29
10.0	1.39	1.44	1.34

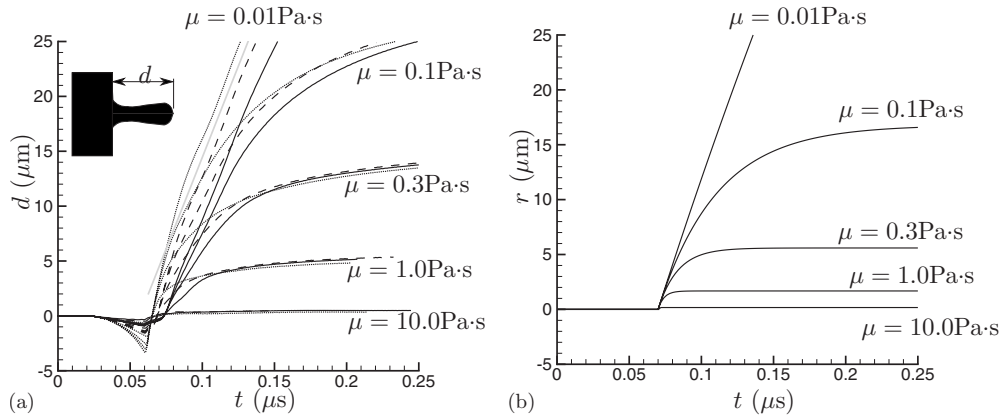


FIG. 5. Penetration distance d into the viscous region. (a) Simulation results for different initial bubble offsets distances: (·····) $w=0$; (----) $w=D/4=5$ μm; and (——) $w=D/2=10$ μm. The different sets of curves are for different viscosities as labeled. The straight gray line is for reference. (b) The penetration model (7) with $t_o=0.7$ μs to facilitate comparison.

indeed nearly inviscid as anticipated based on the Reynolds number estimates in Sec. II.

For the cases with substantial penetration, visualizations such as in Fig. 6 show that the leading portion of the jet has a vortex-ring character. Recognizing that it is roughly spherical and that the jet's fluid is inviscid, we can estimate the drag on the penetrating fluid by the value for an inviscid fluid sphere in a viscous fluid flowing without significant inertia: $4\pi V\mu a$, where a is its radius and V is its speed. We rely on the geometric insensitivity of flow in the Stokes limit in making this estimate. The penetration depth history for $t \geq t_0$ can then be estimated via a solution of the equation of motion as

$$d(t) = V_o \frac{\rho a^2}{3\mu}\left[1 - \exp\left(-\frac{3\mu}{\rho a^2}(t - t_o)\right)\right], \tag{7}$$

where V_o is the initial velocity. This velocity is estimated by identifying the speed of a moving frame of reference that has a stagnation point on the $r=0$ interface between fluid 1 and fluid 2 at the beginning of penetration. For the $\mu=0.3$ Pa s, $w=D/4$ case, this velocity is $V_o=410$ m/s. Given the similarity of the collapses (Figs. 2–4) and of the peak velocities (Table I), we take this V_o for all cases. The sphere radius is taken to be 3.5 μm, based on visualizations. This is, of course, an approximation since it is not exactly the same for all the test cases and can also vary in time for a single case. We also neglect the fact that the vortex-ring "sphere" has a trailing tail of low viscosity fluid and that a significant portion of its trajectory for the higher viscosity cases involves its entry into the viscous fluid, where it should have less drag. Both of these factors would tend to cause Eq. (7) to underpredict penetration depth. We also do not expect perfect numerical agreement since Reynolds numbers are only around unity, and so not in the strictly viscous limit where the $4\pi V_o\mu a$ applies exactly. Despite all these approximations, the predictions from Eq. (7) are remarkably good, as seen in Fig. 5(b). They are within a factor of 2 with those from the simulations in Fig. 5(a), which is as good a comparison as can be expected given the crudeness of this model. Given the translational velocity and diameters of a jet formed by a collapsed bubble, it seems that this model can provide an estimate of penetration depth, trajectory, and perhaps most useful whether or not significant penetration can be expected. It predicts the behavior of both the most viscous and most inviscid cases almost perfectly, though the latter is merely a consequence of the Stokes drag also being near zero for this low value of μ. We could not expect to predict the eventually slowing of this vortex-ring jet given that the drag is not expected to be Stokesian in this case.

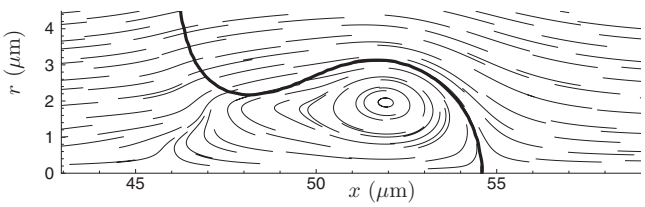


FIG. 6. Streamlines of penetrating jet in frame of reference translating to the right at $U=0.6\sqrt{p0/\rho 0}=190$ m/s for the $\mu=0.3$ Pa s case. The thick line is the $\phi_s=0.5$ isopleth.

When the viscosity is high, the jet obviously does not penetrate the viscous fluid substantially. However, it does cause apparently large shear stresses on it. These simulations were not designed to resolve the thin boundary layers in fluid 1, so we are unable to calculate this stress precisely. The several-mesh-point thickness of the material interfaces in our numerical solution also makes this more challenging. It seems that over an order of magnitude more mesh points in the wall normal direction would be needed to resolve the boundary layers. However, we can estimate the wall rate of strain, which can be seen to be high in Fig. 7. The peak y velocity adjacent fluid 2 is $v_{\max}=866$ m/s. Assuming a linearly decreasing velocity between this peak and the $\phi_s=0.5$ contour with $\mu_1=0.001$ Pa s gives a shear stress estimate $\tau_w=2.3$ MPa. This high level is transient but well over what is needed to cause large deformations and potential injury to cells, which typically have an ~1 kPa Young modulus (e.g., Ref. 44). This "scrubbing" action of the jet when penetration is resisted might explain the apparent damage of the endothelium observed in blood vessels containing cavitation nuclei and subjected to HIFU.[45] The velocity in fluid 2 in this case remains low. Any elasticity would, of course, further resist deformation.

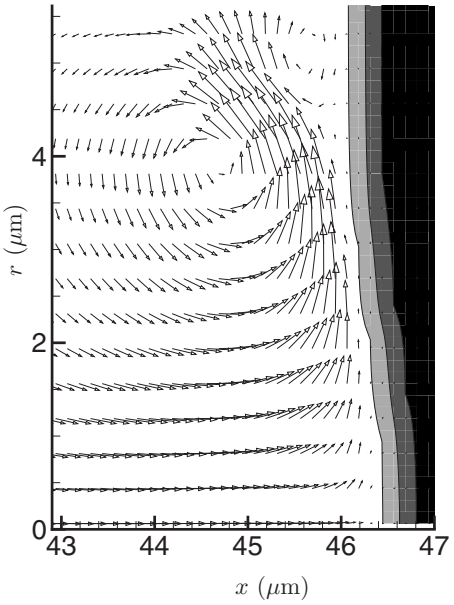


FIG. 7. Velocity vectors showing the high strain rate parallel to the viscous region for the highest viscosity case (μ=10 Pa s) at simulation time t =0.83 μs. Only every third vector is shown in the y direction. The gray levels show the finite thickness phase-field model of the interface between the two fluids. The contour levels are ϕ=0.01,0.5,0.99.

V. A MODEL FOR INJURY SPREAD

A. Injury background

Images of the microstructure of injured kidneys, such as that reproduced in Fig. 8, suggest that injury spreads in sharp fronts behind which the tissue appears utterly disrupted.[46] The tissue in this image was fixed by vascular perfusion immediately following the delivery of 1000 shock waves. This rapid fixing and relatively short treatment time was done so that the primary mechanical shock-wave injury could be distinguished from any subsequent injury due to hypoxia or other mechanisms. The Donier HM-3 used in this study is expected to have a peak shock pressure of about 40 MPa,[26] which matches our shock model. The area of complete disruption in Fig. 8 is bounded approximately with the line added to the image. The tubules on the other side of this line and a few isolated in the disrupted region appear intact, but when examined in greater detail the cells there show blebbing and enlarged electron-clear vacuoles suggestive of localized ischemic injury.[46] The cells and tubules near this borderline also seem to show more injury on sides adjacent to the disrupted area,[46] which is consistent with the apparent spreading character of the injury. Within about three tubules from the disrupted region, the tissue appears completely normal.

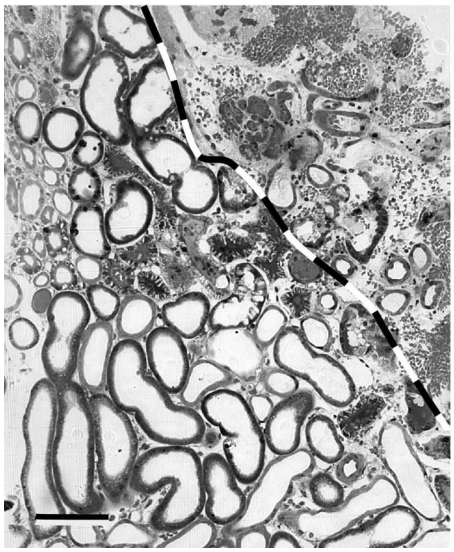


FIG. 8. Image of renal injury similar to that of Shao *et al.* (Ref. 46) with line added demarking the apparently sharp boundary between regions of utterly disrupted (top, right) and intact tubules. The intact tubules immediately adjacent the line drawn show signs of ischemic injury, but within 3 or so tubules distances, they appear completely normal (bottom, left). The scale bar is approximately 50 μm wide.

Our principal concern is the primary mechanical disruption and its spread. We assume that jet penetration causes the disruption and that the mechanical effect of this disruption is to significantly reduce the dissipative resistance of the tissue. Simplistically, the viscosity of any material depends on how it irreversibly transports momentum when transversely sheared. When disrupted at a microscopic level, as we see in Fig. 8, this should be strongly suppressed. In our model this is represented by a drop in viscosity down to a level that offers little resistance to bubble-collapse induced jetting, such as that of fluid 1. For the 20 μm diameter bubbles of the simulations, this would need to be $\lesssim$0.1 Pa s, which is over a factor of 20 higher than the viscosity of whole blood. For larger bubbles, the viscosity of the damaged tissue would need to be proportionally less. Bleeding into this lesion will also reduce the local viscosity toward that of blood, which is not expected to significantly suppress bubble jetting. The numerous small specks in the obviously injured region in Fig. 8 are red blood cells.

From Fig. 8, it appears that the spreading occurs along a front that roughly tracks the tissue microstructure. However, to attempt to relate the general features of the spreading to our jetting model, we assume that the region of injury is spherical with radius $R(t)$. This $R(t)$, of course, should only be regarded as the scale of the injured zone. Many factors are expected to affect its actual shape, including the local microstructure of the kidney and the direction of the shock waves; spherical injury *per se* is not expected and not observed. Indeed, if shock-induced collapse is the root mechanism, we might expect spreading to predominantly occur in the direction of shocks, perhaps leading to roughly a cone shape lesion. Unfortunately, lesion shapes have not yet be quantified sufficiently to make any such assessment of lesion geometry, though this would be interesting and potentially important. The jetting instability of collapsing bubbles is also potentially important, and by its nature might be expected to be more isotropic in its action, perhaps leading to lesions that are roughly spherical. Johnsen and Colonius[20] showed that jet velocities formed by such Rayleigh collapses are not too different from shock-induced collapse, and so should also be governed approximately by Eq. (8). It would be overly ambitious given current understanding to propose a model for the detailed structure of injury. What we hope to see with our single-length-scale model is that Eq. (7) leads to a spreading

that is consistent with the observations in lithotripsy. It should take around 1000 shocks before widespread injury is observed,[2,46] yet upon reaching a typical clinical dose of 2000 shocks the injury should have an ~1 cm scale,[46,47] whatever its actual shape. Lesion geometric effects will be grouped into an undetermined efficiency parameter g.

B. The model

The apparent success of Eq. (7) for penetration depth plus some additional assumptions yields a crude but interesting and potentially relevant description of injury spread through a tissue whose resistance to injury is primarily dissipative. From Eq. (7) we take the depth of injury of a single jet to be

$$\Delta_i = \frac{\rho V_o}{3\mu}\frac{f^2 D^2}{4}, \tag{8}$$

where f is the diameter of the penetration relative to the bubble diameter. This expression follows from Eq. (7) with $t\to\infty$ and $a=fD/2$. Based on our simulations, $f\approx 0.35$. This jet may be shock induced, as in the simulations in this paper, but it also might occur as a bubble is first expanded by the negative-pressure portion of a lithotripter wave and then collapses. We recognize that this penetration distance is approximately independent of bubble stand-off distance [e.g., Fig. 5(a)].

We assume that injury starts in a small region of radius $R(t)=R_o$ and that there are few such regions in the kidney. Adding cavitation nuclei to the systemic circulation is seen to significantly increase injury all over the kidney,[48] which suggests that its initiation under normal conditions is indeed a rare event. The nature of this initial injury under normal circumstances remains unclear, but it has been speculated about for some time. It has been suggested that it might start in some small, slow-flowing vessel in which cavitation nuclei can grow because they are not advected away between shock waves. The expansion of these might initiate injury.[4,7] It might also start at a site of shock-shear induced bleeding.[4,46,49,50] A shear-formed lesion with pooled blood would presumably be more receptive to cavitation injury. Here, we simply assume that this initial region has a radius $R_o\approx 5$ μm, corresponding to a small blood vessel.

The size of the injured region $R(t)$ is assumed to restrict the maximum size of a bubble that may exist there, thus permitting larger bubbles to exist as injury spreads. For large injury zones, however, we assume that finite surface tension or other factors such as the debris left behind by the tissue disruption take over to limit the maximum size of the bubble to some D_m. Pictures of bubbles in water at the target focal point of a lithotripter suggest an upper limit on bubble size of 0.5–1 mm.[4,51] Respecting these two limiting behaviors, we choose to model the bubble diameter by the continuous function

$$D = D_m \tanh\left(\frac{2R}{D_m}\right), \tag{9}$$

and take $D_m=0.5$ mm. An implicit assumption is that after the compression phase of the lithotripter wave passes there is an expansion, which causes the bubbles to re-grow. This is potentially destructive, but based on resistance to expansion estimates,[24] we assume that this action does not extend $R(t)$ when the bubbles are still small. It is also presumed that the violent action of the collapsing and re-expanding bubbles spread bubbles and effectively spawn new nuclei throughout the region of injury. It should be clear that D and D_m in Eq. (9) refer to their size when the next shock comes. Here the bubble is thought to be primarily composed of gas which has come out of solution.[6] At atmospheric pressure, the lifetimes of these bubbles for typical lithotripter shocks is thought to be around 60 s.[4]

Assuming that the volume of the injury increases by the volume of the jet penetration and using Eq. (8), the predicted growth rate of the region of injury is thus

$$\frac{dR}{dt} = S_r g \frac{V_o \rho}{3\mu}\frac{f^2}{4} D_m^2 \tanh^2\left(\frac{2R(t)}{D_m}\right). \tag{10}$$

We take $V_o=410$ m/s based on the jet penetration data from Sec. IV and the shock-wave delivery rate $S_r=1$/s. The viscosity associated with the disruption of tissue at this scale is quite uncertain; following on the discussion in Sec. II we take it to be $\mu=1$ Pa s. Finally, the parameter g is included in Eq. (10) as a model of damage efficiency. It accounts for the finite diameter of the jet in eroding the tissue, anisotropy of the spreading of a non-spherical lesion, and any other factors (e.g., out of phase bubbles with the shock) that mediate the spreading rate. It has long been known with regard to the cavitation damage of surfaces in high-speed liquid flows that only a small fraction (e.g., $g\approx 10^{-4}$) of collapsing bubbles manage to actually damage the surface.[8] Anisotropy of spreading, such as spreading predominantly in the shock direction, will reduce g accordingly. Without the possibility of firmer estimates, we can consider g, or perhaps g/μ together, as a single adjustable parameter to see if we can match any of the phenomenology observed in actual tests with kidney tissue, with the expectation that g will be well less than unity though well more than the $\approx 10^{-4}$ value for flow driven cavitation because the bubbles are confined by the tissue.

To craft Eq. (10), we assume that the injured region expands spherically due to shock induced jetting. In Sec. V A, we recognized that $R(t)$ can only in truth be considered a scale of the region of injury. We should also recognize that shock-induced jetting is not the only potential for spreading injury as the bubbles become larger. Bubble expansion is potentially injurious for larger bubbles and so is the jetting associated with any Rayleigh (non-shock-induced) collapse of expanded bubbles. The basic model we construct here can include the details of the micromechanisms of injury as they are better understood; for now, the principal objective with this model is to show that its basic precepts lead to an apparent injury threshold and injury extent comparable to observation.

An $R(t)$ solution for $g=0.01$ is shown in Fig. 9. The most interesting aspect of this solution is the clear threshold behavior: $R(t)$ remains small for $t\lesssim 1000$ s, which corresponds well with observations from pig kidneys.[2] This switch-over point is particularly sensitive to g. After that point, there is a change to rapidly increasing injury. For these

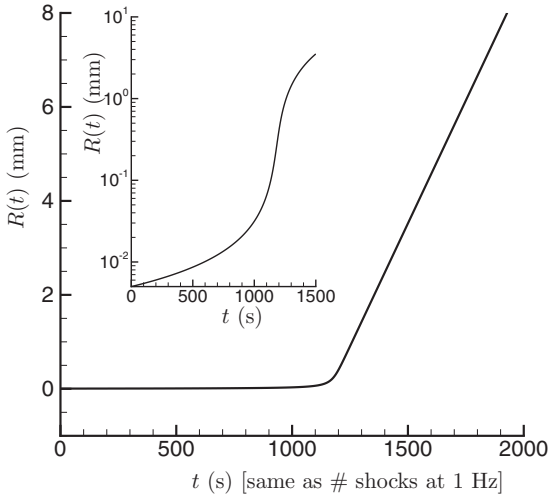


FIG. 9. Model for injury size: a solution of Eq. (10) for parameters discussed in the text.

model parameters, $R(t)$ reaches nearly 1 cm, which is indeed comparable with the lesion size observed in kidney tissue.[46,47] The long-time spreading rate and therefore the lesion size is sensitive to D_m, but it is noteworthy that with a reasonable estimate of D_m from *in vito* lithotriptor bubbles sized, a single g seems to capture both the threshold and extent of injury.

While this crude model does indeed predict behavior that is consistent with observations concerning the threshold behavior and lesion sizes observed, it does not explain the apparent increase of injury with increasing shock-wave delivery rate.[52,53] This rate dependence might be associated with the formation of the initial injury. Increased time between shocks to advect cavitation nuclei in the blood stream away from the focal region or the need for an initial insult via a non-cavitational mechanism might both introduce a rate dependence by delaying the onset of the above spreading mechanism.[50] The bubble diameter D, modeled as time independent in Eq. (9), also has the potential to introduce a rate dependence. It is expected that with sufficient time the bubbles will effectively vanish, though this is expected to be slow at typical conditions.[4] Thus, the expected shock-rate dependence of D or D_m might also introduce rate dependence into the model. Such a dependence could be included by modeling bubble diffusion,[4,6] but the model is probably in need of further testing before building upon it would be fruitful. Tests with erodible materials with known material properties and observable bubble dynamics would be invaluable for assessing this in greater detail.

VI. SUMMARY

In summary, we have shown that viscous resistance of the kind expected in tissue can significantly suppress penetration of bubble-collapse induced jetting. A simple model, which matches these data, was built into a phenomenological model for spreading injury via this mechanism. With one adjustable parameter (g), which we can only anticipate to be well less than unity, this model reproduces the apparent ~1000-shock threshold behavior seen in shock-wave injury of kidneys. For reasonable estimates of the maximum *in vivo* bubble size, the extent of predicted contiguous injury is comparable to that observed.

ACKNOWLEDGMENTS

Portions of this work were supported by NIH Grant No. PO1-DK043881 and the U.S. Department of Energy through the University of California under Subcontract No. B523819.

[1]A. P. Evan and J. A. McAteer, "Q-effects of shock-wave lithotripsy," in *Kidney Stones: Medical and Surgical Management*, edited by F. L. Coe, M. J. Favus, C. Y. C. Pak, J. H. Parks, and G. M. Preminger (Lippincott-Raven, Philadelphia, 1996), pp. 549–570.
[2]A. P. Evan, L. R. Willis, J. E. Lingeman, and J. A. McAteer, "Renal trauma and the risk of long-term complications in shock wave lithotripsy," Nephron **78**, 1–8 (1998).
[3]A. P. Evan and L. R. Willis, "Extracorporeal shock wave lithotripsy: Complications," in *Smith's Textbook on Endourology*, edited by A. D. Smith, G. H. Badlani, D. H. Bagley, R. V. Clayman, S. G. Docimo, G. H. Jordan, L. R. Kavoussi, B. R. Lee, J. E. Lingeman, G. M. Preminger, and J. W. Segura (BC Decker, Inc., Hamilton, Ontario, 2007), pp. 353–365.
[4]M. R. Bailey, L. A. Crum, A. P. Evan, J. A. McAteer, J. C. Williams Jr., O. A. Sapozhnikov, R. O. Cleveland, and T. Colonius, "Cavitation in shock wave lithotripsy," in The Fifth International Symposium on Cavitation, Osaka, Japan (2003), No. Cav03-OS-2-1-006.
[5]M. R. Bailey, Y. A. Pishchalnikov, O. A. Sapozhnikov, R. O. Cleveland, J. A. McAteer, N. A. Miller, I. V. Pishchalnikova, B. A. Connors, L. A. Crum, and A. P. Evan, "Cavitation detection during shock-wave lithotripsy," Ultrasound Med. Biol. **31**, 1245–1256 (2005).
[6]C. C. Church, "A theoretical study of cavitation generated by an extracorporeal shock wave lithotripter," J. Acoust. Soc. Am. **86**, 215–227 (1989).
[7]P. Zhong, I. Cioanta, S. Zhu, F. H. Cocks, and G. M. Preminger, "Effects of tissue contraint on shock wave-induced bubble expansion *in vivo*," J. Acoust. Soc. Am. **104**, 3126–3129 (1998).
[8]T. B. Benjamin and A. T. Ellis, "The collapse of cavitation bubbles and the pressures thereby produced against solid boundaries," Philos. Trans. R. Soc. London, Ser. A **260**, 221–240 (1966).
[9]C. F. Naude and A. T. Ellis, "On the mechanisms of cavitation damage by nonhemispherical cavities collapsing in contact with a solid boundary," ASME J. Basic Eng. **83**, 648–656 (1961).
[10]N. K. Bourne and J. E. Field, "Shock-induced collapse and luminescence by cavities," Philos. Trans. R. Soc. London, Ser. A **357**, 295–311 (1999).
[11]V. K. Kedrinskii, "The role of cavitation effects in the mechanisms of destruction and explosive processes," Shock Waves **7**, 63–76 (1997).
[12]E. Klaseboer, S. W. Fong, C. K. Turangan, B. C. Khoo, A. J. Szeri, M. L. Calvisi, G. N. Sankin, and P. Zhong, "Interaction of lithotripter shockwaves with single inertial cavitation bubbles," J. Fluid Mech. **593**, 33–56 (2007).
[13]Z. Xu, J. B. Fowlkes, and C. A. Cain, "A new strategy to enhance cavitational tissue erosion using a high-intensity, initiating sequence," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 1412–1424 (2006).
[14]B. Sturtevant, "Shock wave physics of lithotriptors," in *Smith's Textbook of Endourology*, edited by A. D. Smith (Quality Medical, St. Louis, MO, 1996), Vol. **1**, pp. 529–552.
[15]E. Klaseboer and B. C. Khoo, "An oscillating bubble near an elastic material," J. Appl. Physiol. **96**, 5808–5818 (2004).
[16]A. Pearson, J. R. Blake, and S. R. Otto, "Jets in bubbles," J. Eng. Math. **48**, 391–412 (2004).
[17]M. Lee, E. Klaseboer, and B. C. Khoo, "On the boundary integral method for the rebounding bubble," J. Fluid Mech. **570**, 407–429 (2007).
[18]E. Johnsen and T. Colonius, "Implementation of WENO schemes in compressible multicomponent flow problems," J. Comput. Phys. **219**, 715–732 (2006).
[19]E. Johnsen and T. Colonius, "Shock-induced collapse of a gas bubble in shockwave lithotripsy," J. Acoust. Soc. Am. **124**, 2011–2020 (2008).
[20]E. Johnsen and T. Colonius, "Numerical simulations of non-spherical bubble collapse," J. Fluid Mech. **629**, 231–262 (2009).
[21]Z. Ding and S. M. Gracewski, "The behaviour of a gas cavity impacted by a weak or strong shock wave," J. Fluid Mech. **309**, 183–209 (1996).
[22]S. Popinet and S. Zaleski, "Bubble collapse near a solid boundary: a numerical study of the influence of viscosity," J. Fluid Mech. **464**, 137–163 (2002).

[23]X. Liu, J. He, J. Lu, and X. Ni, "Effect of liquid viscosity on a liquid jet produced by the collapse of a laser-induced bubble near a rigid boundary," Jpn. J. Appl. Phys. **48**, 016504 (2009).
[24]J. B. Freund, "Suppression of shocked-bubble expansion due to tissue confinement with application to shock-wave lithotripsy," J. Acoust. Soc. Am. **123**, 2867–2874 (2008).
[25]L. Tran and H. S. Udaykumar, "Simulation of void collapse in an energetic material, Part 1: Inert case," J. Propul. Power **22**, 947–958 (2006).
[26]R. O. Cleveland, M. R. Bailey, N. Fineberg, B. Hartenbaum, M. Lokhandwalla, G. A. McAteer, and B. Sturtevant, "Design and characterization of a research electrohydraulic lithotripter patterned after the Dornier HM3," Rev. Sci. Instrum. **71**, 2514–2525 (2000).
[27]S. Girnyk, A. Barannik, E. Barannik, V. Tovstiak, A. Marusenko, and V. Volokhov, "The estimation of elasticity and viscosity of soft tissues in vitro using the data of remote acoustic palpation," Ultrasound Med. Biol. **32**, 211–219 (2006).
[28]E. L. Madsen, H. J. Sathoff, and J. A. Zagzebski, "Ultrasonic shear wave properties of soft tissues and tissue like materials," J. Acoust. Soc. Am. **74**, 1346–1355 (1983).
[29]L. A. Frizzell, E. L. Carstensen, and J. F. Dyro, "Shear properties of mammalian tissues at low megahertz frequencies," J. Acoust. Soc. Am. **60**, 1409–1411 (1977).
[30]S. Nasseri, L. E. Bilston, and N. Phan-Thien, "Viscoelastic properties of pig kidney in shear, experimental results and modelling," Rheol. Acta **41**, 180–192 (2002).
[31]G. N. Sankin, W. N. Simmons, S. L. Zhu, and P. Zhong, "Shock wave interaction with laser-generated single bubbles," Phys. Rev. Lett. **95**, 034501 (2005).
[32]R. Courant and K. O. Friedrichs, *Supersonic Flow and Shock Waves* (Wiley-Interscience, New York, 1948).
[33]A. J. Szeri, B. D. Storey, A. Pearson, and J. R. Blake, "Heat and mass transfer during the violent collapse of nonspherical bubble," Phys. Fluids **15**, 2576–2586 (2003).
[34]J. Rosina, E. Kvasnak, D. Suta, H. Kolarova, J. Malek, and L. Krajci, "Temperature dependence of blood surface tension," Physiol. Res. **56**, S93–S98 (2007).
[35]M. A. Averkioua and R. O. Cleveland, "Modeling of an electrohydraulic lithotripter with the kzk equation," J. Acoust. Soc. Am. **106**, 102–112 (1999).
[36]R. J. LeVeque, *Finite Volume Methods for Hyperbolic Problems* (Cambridge University Press, Cambridge, 2002).
[37]E. F. Toro, M. Spruce, and W. Speares, "Restoration of the contact surface in the HLL-Riemann solver," Shock Waves **4**, 25–34 (1994).
[38]E. Olsson, G. Kreiss, and S. Zahedi, "A conservative level set method for two phase flow II," J. Comput. Phys. **225**, 785–807 (2007).
[39]Y. Sun and C. Beckermann, "Sharp interface tracking using the phase-field equation," J. Comput. Phys. **220**, 626–653 (2007).
[40]R. K. Shukla, C. Pantano, and J. B. Freund, "An interface capturing method for simulation of multiphase compressible flows," J. Comput. Phys. submitted (2009).
[41]C. A. Kennedy and M. H. Carpenter, "Additive Runge-Kutta schemes for convection-diffusion-reaction equations," Appl. Numer. Math. **44**, 139–181 (2003).
[42]H. A. van der Vorst, "Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems," SIAM (Soc. Ind. Appl. Math.) J. Sci. Stat. Comput. **13**, 631–644 (1992).
[43]C. K. Turangan, A. R. Jamaluddin, G. J. Ball, and T. G. Leighton, "Free-Lagrange simulations of the expansion and jetting collapse of air bubbles in water," J. Fluid Mech. **598**, 1–25 (2008).
[44]H. Karcher, J. Lammerding, H. Huang, R. T. Lee, R. D. Kamm, and M. Kaazempur-Mofrad, "A three-dimensional viscoelastic model for cell deformation with experimental verification," Biophys. J. **85**, 3336–3349 (2003).
[45]J. H. Hwang, J. Tu, A. A. Brayman, T. J. Matula, and L. A. Crum, "Correlation between inertial cavitation dose and endothelial cell damage *in vivo*," Ultrasound Med. Biol. **32**, 1611–1619 (2006).
[46]Y. Shao, B. A. Connors, A. P. Evan, L. R. Willis, D. A. Lifshitz, and J. E. Lingeman, "Morphological changes induced in the pig kidney by extracorporeal shock wave lithotripsy," Anat. Rec. **275A**, 979–989 (2003).
[47]P. M. Blomgren, B. A. Connors, J. E. Lingeman, L. R. Willis, and A. P. Evan, "Quantitation of shock wave lithotripsy-induced lesion in small and large pig kidneys," Anat. Rec. **249**, 341–348 (1997).
[48]B. R. Matlaga, J. A. McAteer, B. A. Conors, R. K. Handa, A. P. Evan, J. C. Williams, J. E. Lingeman, and L. R. Willis, "Potential for cavitation-mediated tissue damage in shockwave lithotripsy," J. Endourol **22**, 121–126 (2008).
[49]M. Lokhandwalla and B. Sturtevant, "Mechanical haemolysis in shock wave lithotripsy (SWL): I. Analysis of cell deformation due to SWL flowfields," Phys. Med. Biol. **46**, 413–437 (2001).
[50]J. B. Freund, T. Colonius, and A. P. Evan, "A cumulative shear mechanism for tissue damage initiation in shock-wave lithotripsy," Ultrasound Med. Biol. **33**, 1495–1503 (2007).
[51]Y. A. Pishchalnikov, J. A. McAteer, and J. C. Williams, "Effect of firing rate on the performance of shock wave lithotriptors," BJU Int. **102**, 1681–1686 (2008).
[52]J. A. McAteer, A. P. Evan, J. C. Williams, Jr., and J. E. Lingeman, "Treatment protocols to reduce renal injury during shock wave lithotripsy," Curr. Opin. Neurol. **19**, 192–195 (2009).
[53]A. Evan, J. A. McAteer, B. A. Connors, P. M. Blomgren, and J. E. Lingeman, "Renal injury during shock wave lithotripsy is significantly reduced by slowing the rate of shock wave delivery," BJU Int. **100**, 624–628 (2007).

Acoustic radiation patterns of mating calls of the túngara frog (*Physalaemus pustuosus*): Implications for multiple receivers

Ximena E. Bernal[a] and Rachel A. Page[b]
Section of Integrative Biology, University of Texas at Austin, 1 University Station C0930, Austin, Texas 78712

Michael J. Ryan
Section of Integrative Biology, University of Texas at Austin, 1 University Station C0930, Austin, Texas 78712 and Smithsonian Tropical Research Institute, P.O. Box 0943-03092, Balboa Ancón, Republic of Panamá

Theodore F. Argo IV and Preston S. Wilson
Department of Mechanical Engineering and Applied Research Laboratories, University of Texas at Austin, P.O. Box 8029, Austin, Texas 78713-8029

(Received 22 November 2008; revised 30 July 2009; accepted 31 July 2009)

In order for a signal to be transmitted from a sender to a receiver, the receiver must be within the active space of the signal. If patterns of sound radiation are not omnidirectional, the position as well as the distance of the receiver relative to the sender is critical. In previous measurements of the horizontal directivity of mating calls of frogs, the signals were analyzed using peak or root-mean-square analysis and resulted in broadband directivities that ranged from negligible to a maximum of approximately 5 dB. Idealized laboratory measurements of the patterns of acoustic radiation of the mating calls of male túngara frogs (*Physalaemus pustulosus*), along axes relevant to three receivers in this communication network, female frogs in the horizontal plane, and frog-eating bats and blood-sucking flies above the ground, are reported. The highest sound pressure level was radiated directly above the frog, with a 6 dB reduction radiated along the horizontal direction. Band-limited directivities were significantly greater than broadband directivities, with a maximum directivity of 20 dB in the vertical plane for harmonics near 6 kHz. The implications with regard to mating and predator-prey interactions are discussed.
 [DOI: 10.1121/1.3212929]

PACS number(s): 43.80.Ka [MJO] Pages: 2757–2767

I. INTRODUCTION

In its simplest form, communication is a dyadic interaction between a signaler and a receiver in which the signal has some probabilistic influence on the behavior of the receiver.[1] For communication to proceed, the signal must be detected and perceived by the receiver; that is, the receiver must be within the active space of the signal. In acoustic communication, the size of the active space is dependent on the amplitude of the signal at the source, the patterns of radiation of the signal, and the sensitivity of the receiver. If patterns of sound radiation are not omnidirectional, the position of the receiver relative to the sender is critical. In many animal systems, the radiation of acoustic signals is directional, often with higher amplitude anterior to the sender with a bilaterally symmetric sound field around it.[2–5] Some species of frogs and toads produce nearly omnidirectional acoustic radiation in the horizontal plane while others have 5 dB or more of directivity.[6] In some species, such as the sage grouse, the directionality or beam pattern of the sound is more pronounced.[7]

Not all communication is dyadic. In many systems animals send signals to more than one receiver within a social group of conspecifics. Quantifying directionality of the source is important to define the communication network. In many if not most acoustic communication systems, such as chorusing insects, frogs, and birds, the multiple conspecific receivers are often in the same plane and have similar thresholds for signal detection. Quite often, however, there are unintended receivers or "eavesdroppers."[8,9] These receivers attend to the same signals as do the conspecifics but for a different reason, they use the signals as acoustic beacons to lead them to potential prey or hosts. In a classic example, the mating calls of male field crickets attract both female crickets for mating and the fly *Ormia*, which locates the call of the cricket and deposits its larvae on the male. The larvae burrow into the male cricket and use him as a food source as they develop.[10] Heterospecifics that eavesdrop on the mating signals of their hosts or prey are widespread across taxonomic groups (e.g., gecko-cricket,[11] bat-katydid,[12] emblysoma fly-cicada,[13] orminiie fly-bushcricket,[14,15] heron-cricket,[16] opossum-frog,[17] and turtle-frog[18]). In cases

[a]Present address: Department of Biological Science, Texas Tech University, Box 43131 Lubbock, TX 79409.
[b]Present address: Sensory Ecology Group, Max Planck Institute for Ornithology, 82319 Seewiesen, Germany.

in which the signaler and the intended receiver communicate in the horizontal plane while the unintended receiver detects the signal in a vertical plane, the characteristics of the signal available to the different receivers may vary greatly. Thus, beaming patterns of the signal influence its effectiveness at attracting mates and the costs imposed by acoustically orienting predators and parasites. Although considerable attention has been devoted to signal adaptations that increase signal transmission through the environment,[19–21] the role of the beam pattern in signal evolution has been largely ignored.

In this study, patterns of acoustic radiation of mating calls of male túngara frogs (*Physalaemus pustulosus*) were measured. This neotropical frog is a well-known model organism for studies of vertebrate communication (see reviews in Refs. 22 and 23). Males produce acoustic signals, or mating calls, that are the primary cue females use to locate and assess males for mating. All calls contain a multi-harmonic frequency sweep, the whine. During the sweep, the first harmonic traverses frequencies from approximately 900 to 400 Hz in approximately 300 ms. The whine can be produced by itself or can be followed by 1–7 short, broadband bursts of sound, the chucks, each with a duration of about 45 ms.[24] The whine by itself, the simple call, is necessary and sufficient to elicit phonotactic responses from females, while the addition of chucks, which form complex calls, increases the attractiveness of the call to females. Males tend to produce simple calls when calling in isolation but escalate to complex calls in response to calls of other males. In this study, few males produced chucks, and, therefore, the chucks were omitted from the analysis.

The production of complex calls is favored by sexual selection because it increases the males' probability of mating. There are, however, two primary eavesdroppers in this system: the frog-eating bat *Trachops cirrhosus*[25] and the blood-sucking fly *Corethrella*,[26] both of which are attracted to the calls of male túngara frogs. Both eavesdroppers have call preferences similar to those of female túngara frogs; they are attracted to simple calls but prefer complex ones.[27] Both eavesdroppers approach the calling males from above while the female frogs approach the males in the horizontal plane.

The purpose of this study is to document in an idealized environment the patterns of acoustic radiation along axes relevant to the three known receivers in this communication network: the female frogs in the horizontal plane, and the frog-eating bats and blood-sucking flies above the ground. The purpose of this study is not to duplicate conditions in which frogs call in nature. In fact, there is no single calling condition because weather, topography, and intervening vegetation at calling sites all vary substantially across time and space. Instead, the purpose is to define a benchmark in which acoustic radiation is quantified precisely with variables eliminated, e.g., intervening vegetation, or controlled, e.g., topography, temperature, and humidity. This benchmark can then be used as a standard against which to assess radiation patterns in the wild.

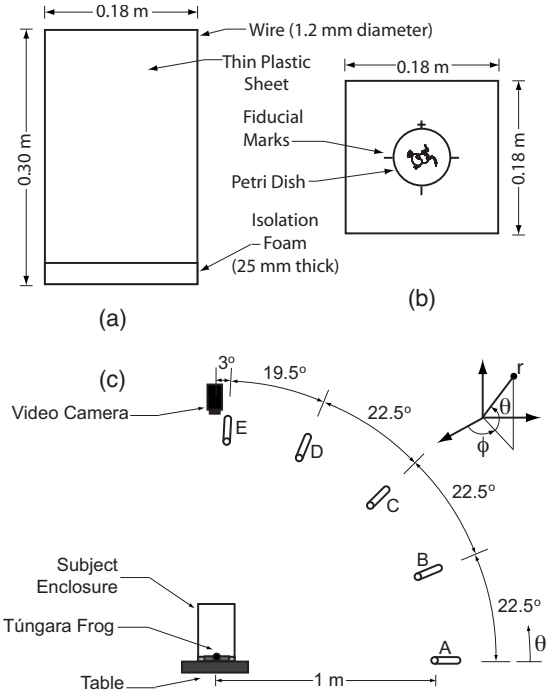


FIG. 1. Diagram of the subject enclosure. Panel (a) shows the side view and panel (b) shows the top view of the enclosure. A schematic (c) of the directivity microphone array is shown. All microphones were placed 1 m from the center of the petri dish. The angular coordinate system is also shown. The angles θ and ϕ are referred to as elevation and azimuth, respectively.

II. MATERIALS AND METHODS

A. Subjects

Ten male túngara frogs were tested from a breeding colony at the University of Texas at Austin, TX from December 8, 2006 to January 5, 2007. Colony frogs were maintained on an adjusted light/dark cycle such that dawn began at 02:00 and dusk began at 14:00. Males were tested from 18:00 to 01:00, during their active period. The mean mass of the males tested was 1.31 g, and the mean snout-vent length was 26.46 mm. These measurements are within the range of measurements of male frogs found in the wild.[22] After testing, the males were returned to the colony and marked using a toe-clipping system to avoid using the same individual more than once in the experiment.

To stimulate calling behavior, males were injected with 500 IU human chorionic gonadotropin (HCG) 24–48 h prior to testing. HCG has been shown to stimulate reproductive behavior in anurans.[28] HCG was dissolved in 0.9% saline solution and injected subcutaneously in a volume of 0.5 ml. In túngara frogs, HCG injection does not alter the characteristics of the call (M. Ryan, personal observation). All tests were licensed and approved by the University of Texas at Austin (IACUC Protocol No. 6041701).

Males were placed one at a time in a $18\times18\times30$ cm^3 enclosure depicted in Fig. 1. The walls of the enclosure consisted of transparent plastic film (thickness=0.0381 mm) loosely supported by 1.2 mm diameter wire. Such an enclosure was previously shown to be acoustically transparent to túngara frog calls.[29] In addition, the acoustic pressure level transmitted by a directional 38 kHz source was measured at

various angles from within the enclosure and compared to the level transmitted in absence of the enclosure. Variation among these measurements was less than 0.5 dB, which is small compared to the directional variation discussed in Sec. III and was thus ignored. The enclosure was open at the top, allowing for unobstructed view from above. The position of the frogs was tracked with a video camera positioned directly above. The base of the enclosure consisted of 25.5 mm thick foam, with a 6.7 cm diameter petri dish inset into the center. The petri dish was filled with water to afford the males an appropriate environment for calling. In nature, túngara frogs are found calling only in shallow water, and it is thought that water is necessary for full expansion of the vocal sac.[22]

The enclosure was positioned in the center of a 30×60 cm^2 table, which itself was placed at the center of a $3.62 \times 2.46 \times 2.20$ m^3 fully anechoic chamber. The table was used to mimic the acoustical effect of the water surface from which túngara frogs call. To waves incident in air, both water and the hard surface of a wood table appear acoustically rigid, and both are smooth and flat. The large walk-in anechoic chamber used in this work is located in the basement of its building, attached to the foundation. It is fully enclosed by an outer shell of solid concrete blocks and an inner shell, also made of solid concrete blocks. There is a 1 ft air gap separating the outer and inner shells on all sides and the inner shell is suspended on springs and dampers for isolation from low frequency structure-borne sound and vibration. Acoustic isolation doors on each shell allow access into the chamber. The inner walls, door, ceiling, and floor of the chamber are fully covered with 3 ft long sound-absorbent fiberglass wedges. The wedges are attached to the inside of the inner shell via a compliant mounting and are placed in groups of three. The orientation of the edges of each group of wedges alternates between horizontal and vertical between neighboring groups. A removable wire mesh platform is suspended above the floor-mounted wedges to allow users to walk into the chamber. Measurements provided by the manufacturer indicate that the noise floor of the chamber is 0–10 dB re 20 μPa, depending primarily on the traffic level on the street outside the building, and that free field conditions exist within the chamber at frequencies above 200 Hz.

The temperature in the entire chamber was controlled with a space heater to obtain temperatures appropriate for the frogs to call (approximately 26 °C). A room humidifier was used to increase the humidity of the air in the chamber, also to facilitate calling by the frogs. A relative humidity of 40%–50% was the maximum that could be achieved. The frog's natural environment usually has a higher humidity, but this humidity difference results in a negligible difference in acoustic propagation. The sound speed change is less than 0.3% for air at 50% and 100% relative humidities, at the experimental temperature,[30] and there is a maximum of 0.03 dB difference in attenuation along the experimental propagation path for this humidity difference.[31] The subjects acclimated for over 1 h inside the enclosure in the chamber before measurements began.

A recording of a túngara frog chorus was then broadcast from a small loudspeaker located approximately 1.5 m from the frog enclosure to evoke calling by the test male. Once the test male began calling in response to the chorus, the amplitude of the chorus playback was gradually reduced until the male called in silence. The spectral content, amplitude, and repetition rate of the calls recorded in the present experiment are typical of calls recorded in the natural environment.[22]

B. Measurement instrumentation

Calling behavior was recorded with a night vision video camera positioned 1 m above the frog enclosure. These data allowed us to determine the orientation and position of the frog within the enclosure for each recorded call. All trials were conducted in darkness, illuminated only by the infrared light on the video camera. Optomotor studies show that túngara frogs are not sensitive to light in the infrared (X. Bernal and M. J. Ryan, unpublished data). The frog's orientation relative to the microphone array was determined by measuring the angle between the centerline connecting the frog's snout to the frog's vent and orthogonal tape marks on the foam base that were aligned with the microphone array.

Wide-bandwidth acoustic pressure recordings (10 Hz–51.4 kHz) of five frogs were obtained with a GRAS model 40BF free-field microphone positioned 0.63 m from the center of the petri dish. Some frogs produce calls in the ultrasonic range,[32] thus measurements were made in this frequency range to document the presence or absence of these frequencies in this species. The microphone was supported by a tripod-style microphone stand and was calibrated by the manufacturer. Its response was flat within ± 1.5 dB from 10 Hz to 50 kHz, and flat within ± 3 dB from 50 to 100 kHz. The microphone cartridge was connected to a GRAS type 26 preamplifier that possessed a flat (± 0.2 dB) bandwidth from 2 Hz to 200 kHz. The signals were digitized with a personal computer based data-acquisition board and a sampling rate of 102.8 kHz.

Audio-band acoustic pressure recordings (10 Hz–22.0 kHz) of five different frogs were obtained using five Sennheiser model ME66 audio-bandwidth microphones placed 1 m from the center of the petri dish in a plane perpendicular to the plane of the table, as shown in Fig. 1(c). The microphones at 0°, 22.5°, and 45° were placed in tripod-style microphone stands and the microphones at 67.5° and 87° were suspended using thin woolen string. The five Sennheiser microphones were calibrated by comparison with one of the GRAS microphones to within 0.2 dB of the GRAS response, which is significantly less than the directivity observed in the measured beam patterns reported in this work. The signals were recorded with a Racal Storeplex multichannel digital tape deck with a 96 dB dynamic range, using a sampling rate of 45.5 kHz. In-line impedance-matching microphone transformers were used to connect the balanced low-impedance microphones to the high-impedance, ground-referenced single-ended inputs on the tape deck. The microphone signals were played from tape and digitized with a data-acquisition system (also 96 dB dynamic range) running on a desktop computer.

C. Signal processing

Signal processing of the ultrasonic bandwidth data consisted of calculating fast Fourier transforms (FFTs) and spectrograms using commercially available signal processing software. For the audio-bandwidth data, commercially available signal processing software was also used to perform the following operations. The remaining discussion in this section applies only to the audio band data. Each channel was detrended to remove any dc-voltage bias. Each frog produced multiple calls in succession; therefore, time gates were applied to isolate single calls. The maximum frequency was typically less than 10 kHz; hence the data were downsampled to 22 kHz. The broadband sound pressure levels (SPLs) were computed for each channel. These SPLs were referenced to the maximum SPL received for that call. The broadband directivity was visualized by plotting on a polar plot the SPL of each channel as a function of the angle at which it was recorded.

Spectrograms were then computed via the short time Fourier transform. The calls were time gated into blocks 512 points in length with a 92.8% overlap (475 points) with the previous block. A 500-point Kaiser window with a beta value of 5 was applied to each block. A 2048-point FFT operated upon each windowed block in succession to produce a spectrogram.

Frequency-dependent directivity for each audio-band call was determined at a particular time t_0 near the beginning of the call, where the highest frequencies were found and where subharmonics were not present, by extracting the FFT at t_0 within the spectrogram. The relationship among the peak frequencies found within the FFT was examined to determine the harmonic structure of the call. Each channel's FFT contained a series of harmonics, the magnitudes of which were extracted using a peak-finding algorithm. The magnitude of each peak for each channel was converted to decibels normalized by the maximum SPL at that peak. Directivity patterns were constructed from the normalized SPLs for each harmonic. The number of harmonics at time t_0 in most calls received at most directions varied between 6 and 8, although 9 harmonics were visible in some signals.

III. RESULTS AND DISCUSSION

Of the ten males tested, five were recorded with the wide-bandwidth GRAS microphone to investigate high-frequency call components, and five were recorded with the audio-frequency-range Sennheiser microphone array to investigate beaming patterns. A total of 66 calls from five males were recorded with the wide-bandwidth GRAS microphone. The spectrogram of a typical call recorded with the wide-bandwidth system is shown in Fig. 2(a). The maximum frequency component that appears above the noise floor is at about 11.5 kHz. Individual FFTs from several times within the spectrogram are shown in Fig. 2(b). The thick black spectrum is close to the noise floor, from a quiet time past the end of the call. The three other spectra are from near the beginning of the call. Peaks that are lower in frequency than the peak labeled (α) are persistent over time. Peak (β) appears by itself. The thin black spectrum and the blue spectrum do not have corresponding peaks at this frequency. The peaks that are higher in frequency than (β) do not persist over time. Therefore, we conclude that the highest frequency that appears in the call is about 11.5 kHz. None of the remaining 65 calls that were recorded with the wideband system contained higher frequency content above the noise floor. This result is consistent with the observation of increasing attenuation above a few kilohertz in the frog's natural environment due to interaction with vegetation.[33]

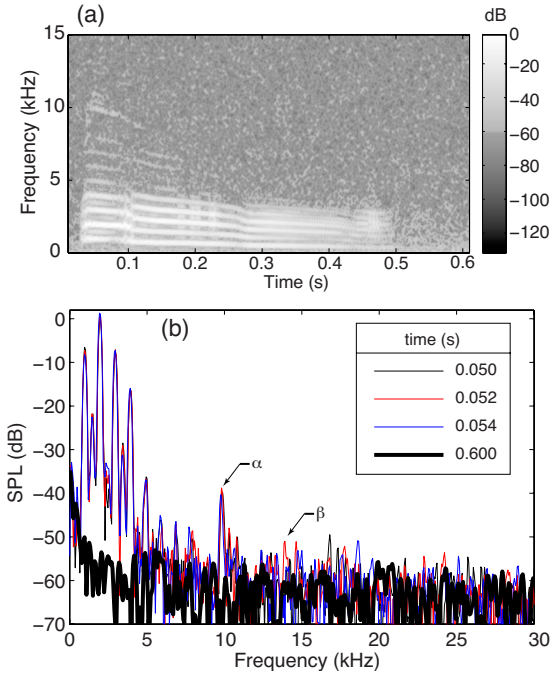


FIG. 2. (Color online) A spectrogram of a typical call recorded with the wide-bandwidth system is shown in (a). The highest persistent frequency component that appears above the noise occurs at 11.5 kHz. FFTs are shown in (b) from four times within the spectrogram of (a). The thick black spectrum is close to the noise floor, past the end of the call. The three remaining spectra are from near the beginning of the call. Peaks lower in frequency than the peak labeled (α) are persistent over time and correspond to the call. Peak (β) appears by itself. The thin black spectrum and the blue spectrum do not have corresponding peaks at this frequency. Peaks higher in frequency than (β) are not persistent over time but vary randomly; hence they are considered noise.

The remainder of the results reported here were obtained with the audio-frequency-range Sennheiser microphone array. Approximately 140 calls from five males were analyzed. A typical call is shown in Fig. 3(a). The waveforms were recorded at each of the five azimuthal angles given in Fig. 1(c). Microphone E is directly above the frog and microphone A is on ground level. A high-amplitude burst is visible at the onset of each waveform, followed by a decay; yet each waveform has a different envelope. For example, at 0.225 s there is a pronounced amplitude reduction in the high angle recordings (C, D, and E) and relatively less amplitude reduction in the low-angle recordings (A and B). In general, signal amplitude is retained at larger angles for a greater amount of time than at lower angles.

The broadband directivity of the call is shown in Fig. 3(b). The SPL of the waveform recorded at the ith angle was calculated with

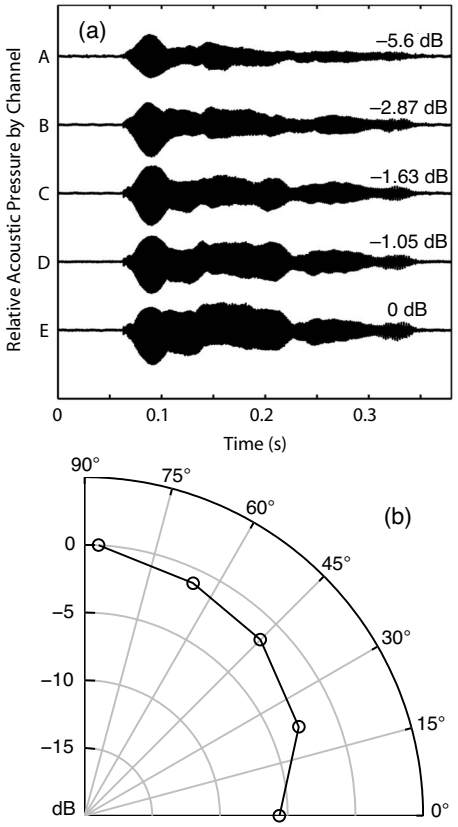


FIG. 3. Time waveforms of a typical túngara frog call recorded with the microphone array are shown in (a). The SPL, relative to the root-mean-square pressure recorded at microphone E (directly above the frog), is calculated for each channel. The broadband directivity in elevation plane θ is presented in (b), using the SPLs shown in (a).

$$\mathrm{SPL}_i = 20 \log_{10}\left[\frac{p_{\mathrm{rms},i}}{p_{\mathrm{rms,E}}}\right] (\mathrm{dB}), \tag{1}$$

where $p_{\mathrm{rms},i}$ is the rms pressure of the waveform recorded at the ith angle, $p_{\mathrm{rms,E}}$ is the rms pressure of highest amplitude waveform (microphone E), and the units are decibels. The highest SPL was radiated directly above the frog and the SPL is reduced at each angle until there is about a 6 dB reduction radiated along the horizontal direction.

Narrow-band directivity was also investigated. A spectrogram of signal E from Fig. 3(a) is shown in Fig. 4(a), where the call is seen to consist of a downward-sweeping chirp. At any given time, the call is composed of a series of harmonics, and the fundamental frequency decreases as time increases. This characteristic pattern of harmonics is shown in Fig. 4(b), for time t_0 indicated by the black vertical line in Fig. 4(a), but the spectra recorded at all five angles are shown. A close-up of the peaks associated with the second harmonic is shown in the inset, Fig. 4(c), where it can be seen that the narrowband amplitude received at each angle is different, and hence there is narrowband directivity, in addition to the broadband directivity already illustrated. Beam patterns are formed using these data in Figs. 5 and 6. The fundamental frequency of the calls in the dataset varied by at most a few percent with individual and from call-to-call in the same individual. Because of this variation, it was convenient to compare narrowband levels as a function of the harmonic number, instead of comparing them directly as a function of frequency.

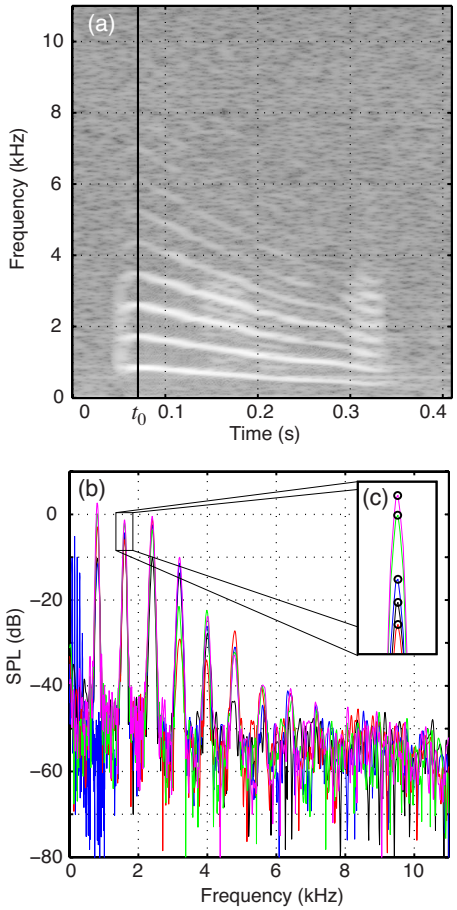


FIG. 4. (Color online) In this spectrogram (a) of a typical call, lighter shades of gray indicate higher amplitude. The time that corresponds to the highest frequency is indicated by the vertical line, at approximately 0.075 s. The FFTs at that time and all angles are shown in (b). At frequencies below 3 kHz peaks rise up to 50 dB above the noise, whereas at frequencies approaching 8 kHz the peaks become indistinguishable from the noise. In (c), the relative amplitude received at different elevation angles θ is shown for the second harmonic. These narrowband SPLs are presented (in Fig. 5) in the form of directivity plots for each harmonic, and for various azimuthal angles ϕ using calls from the same frog. The data in Fig. 5 were all taken at times within the call that corresponded to the highest frequency, as illustrated by the solid line at t_0 in Fig. 4(a).

Nonlinear phenomena are exhibited in the recorded calls. Subharmonics are visible in both the spectrograms and FFTs shown in Figs. 2 and 4. There are also frequency jumps in Fig. 2(a) located at about 0.1 s and just before 0.3 s. Such nonlinear features are common in other species' vocal production mechanisms[34] and calls that contain subharmonics have been documented for the túngara frog.[24] In túngara, these nonlinear features appear to be caused by nonlinear mechanical dynamics of the frog's vocal production mechanism. Specifically, a fibrous mass attached to the vocal folds that can undergo impact oscillation at a sufficiently high excitation level appears to be responsible for the presence of subharmonics in the portion of the call known as the chuck.[24] Subharmonic generation by impact oscillation (also known as clapping or impact nonlinearity) has been documented in many dynamic systems.[35–37]

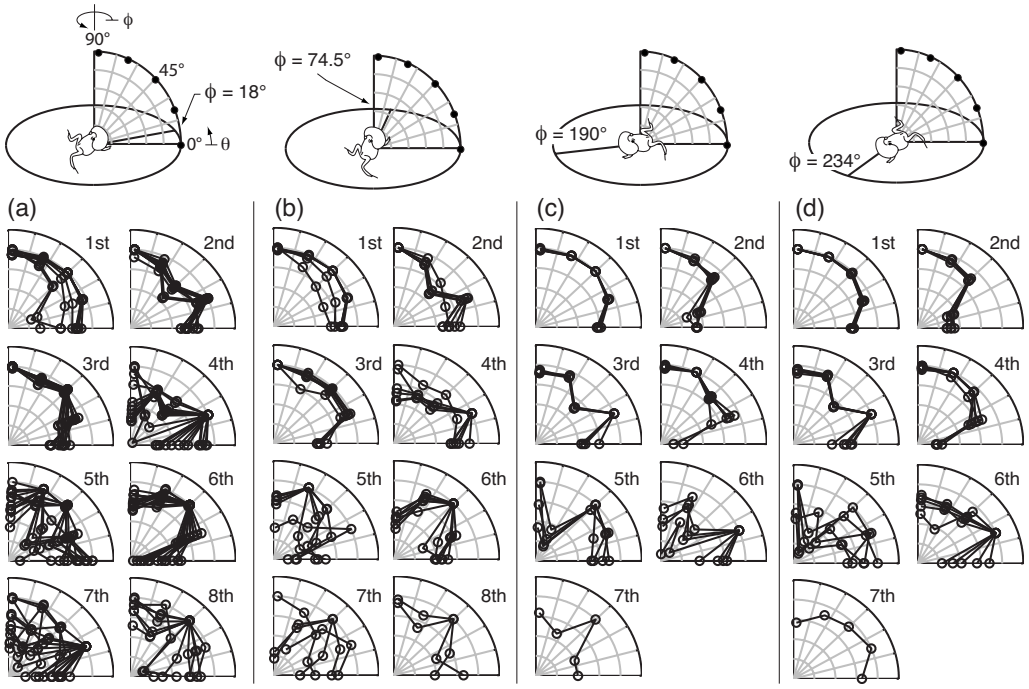


FIG. 5. Narrowband elevation directivity plots for a single frog at several azimuth angles are shown. Thirteen calls recorded at ϕ=18° are shown in (a). Six calls at ϕ=74.5° are shown in (b). Four calls at ϕ=190° are shown in (c). Five calls at ϕ=234° are shown in (d).

Despite these nonlinear phenomena, nonlinear acoustic propagation does not play a role in this work. The SPL of typical túngara frog calls (about 75–85 dB re 20 μPa),[22] and the propagation distance in this work (1 m) indicates that the acoustic propagation is linear. Nonlinear acoustic propagation effects only become important at higher amplitudes and for greater propagation distances, for example, 120 dB re 20 μPa and 100 m, as shown in Fig. 16.3.1 of Ref. 38. We therefore conclude that the sound radiation from the frog, the subsequent propagation, and call directivity are due to linear acoustic diffraction and are not effected by the source production mechanism's nonlinearity.

Directivity in the calls of one individual is illustrated for each harmonic at each of four azimuthal angles in Fig. 5. Beam patterns for several calls are shown in each frame where available. The data clearly display two characteristics. There is significant directivity in many of the harmonics, and there is significant variability from call-to-call, across different azimuthal angles and across different harmonics. The first harmonic generally mimics the broadband directivity of Fig. 3(b), with the main beam pointing directly above the frog, but in two calls in the first harmonic frame in Fig. 5(a), a dipole pattern is present. At higher harmonics, however, fairly strong beams appear, as in the sixth harmonic of Figs. 5(a) and 5(b), where the beam points at 45° above the horizon, as much as 20 dB higher in amplitude than signals oriented along the horizon at 0°. There is a large amount of call-to-call variability in some harmonics, the seventh har-

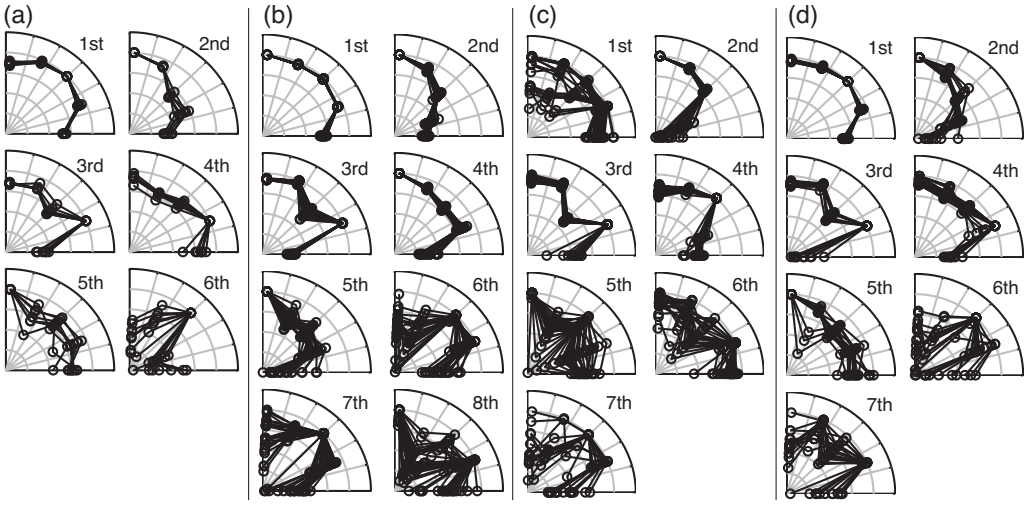


FIG. 6. Narrowband elevation directivity exhibited by four frogs at various azimuth angles is shown. Six calls by Frog 18 at ϕ=177° are shown in (a). Nineteen calls by Frog 19 at ϕ=121° are shown in (b). Twenty-five calls at ϕ=86° are shown in (c). Thirteen calls at ϕ=135° are shown in (d).

monic of Fig. 5(a), for example, while at the same time the third harmonic of Fig. 5(a) shows significantly less variability.

It is also interesting to consider the variability of the beam pattern at a particular harmonic as a function of azimuthal angle. Comparing the third and fourth harmonics in Figs. 5(a) and 5(b), to those in Figs. 5(c) and 5(d) reveals that, in each case, the patterns are different from front to back. For example, in Figs. 5(a) and 5(b), the third harmonic has a peak near 45°, which points up in the anterior direction, but in Figs. 5(c) and 5(d), there is a local minimum at 45°, in the posterior direction. The beams are reversed for the fourth harmonic, where a local minimum occurs in the anterior direction [Figs. 5(a) and 5(b)] and the directivity is relatively flat at 45° in the posterior direction [Figs. 5(c) and 5(d)]. Despite the variability, robust directivity clearly exists on average. The calls could be perceived differently to a listener depending on the relative position. This directivity could play a discrimination role for both intended and unintended listeners. Finally, when taken as a whole using the broadband directivity as a measure [Fig. 3(b)], more energy is directed upward, where the unintended listeners reside—the predators and parasites.[25,26] Relatively less energy is directed along the horizontal plane, where the intended listeners, the females, reside. This condition yields asymmetry between the costs and benefits associated with the túngara frog mating call.

The intention of this work is to illustrate the presence of directivity and variability in túngara calls. This work does not attempt to provide a species-wide generalized description of the call, nor to fully explain the ramifications of the directivity. Nonetheless, the results shown in Fig. 5 for a single individual are typical of the calls made by other males at other azimuthal angles, as illustrated in Fig. 6. At the current stage, there are not enough data from any one individual to fully populate the azimuthal angle parameter space, and not enough data from different individuals to calculate global mean beam patterns at even one angle. The current data do support the two main points mentioned previously: There is significant directivity in túngara frog calls, and the directivity exhibits significant variability from call-to-call, from harmonic-to-harmonic, and from individual-to-individual.

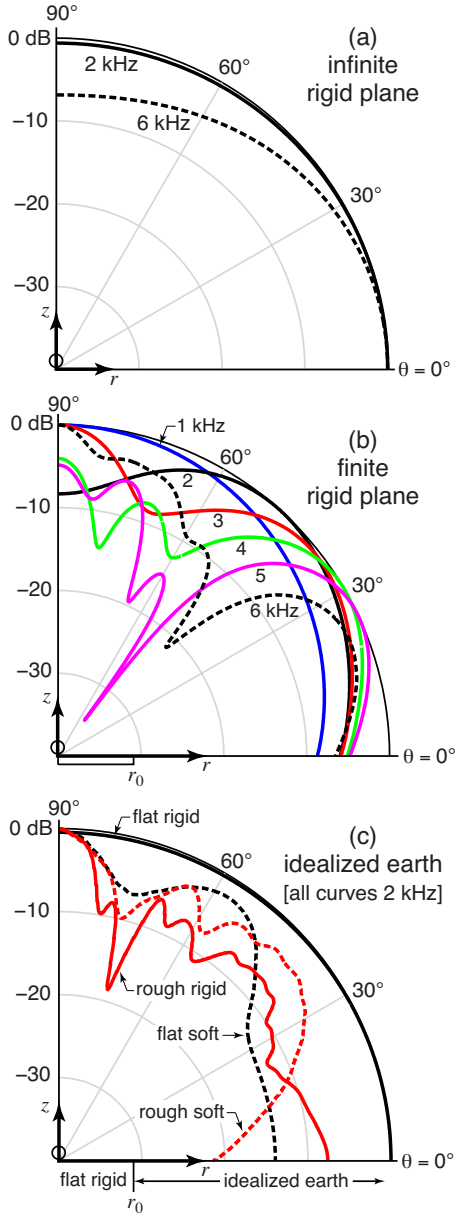


FIG. 7. (Color online) Model geometry and the results of directivity calculations are shown. The source location is indicated by the open circle at $r=0$ and $z=1$ cm. The directivity of a simple source above a rigid plane of infinite extent is shown in (a). The infinite plane lies along the r-axis and is perpendicular to the z-axis. Directivity at 2 and 6 kHz is shown. Directivity due to a finite-sized rigid reflecting plane is shown in (b) for a range of frequencies. The extent of the reflector is indicated by r_0. Directivities due to various representations of an idealized natural environment are shown in (c). A flat rigid reflector resides along the r-axis with radius r_0. Beyond r_0, four different realizations are shown: a continuation of the flat rigid reflector, a flat soft layer, a rough rigid layer, and a rough soft layer. Additional details are in the text.

IV. MODELING OF RADIATION PATTERNS

Several mathematical and numerical models were developed and used to interpret the radiation patterns presented in Sec. III. The goal of this modeling effort was to illuminate the leading order parameters that govern some of the features observed in the vertical plane directivity. The modeling was not intended to explain fine structure or details of either vertical or horizontal directivity. These models are based on the assumption that the acoustically active part of the frog is small compared to the acoustic wavelength for the frequencies discussed here, and hence the frog was modeled as a simple source. All simple sources produce the same acoustic field, that of a uniformly pulsating sphere, regardless of their shape.[38] The vocal sac of the túngara frog is the primary source of acoustic radiation[39–43] and it is approximately spheroidal in shape, with a nominal width during the whine of about 2 cm, for calls without chucks.[41] At the lowest frequencies analyzed in this work, about 1 kHz, the wavelength is about 34 cm, or 17 times the width of the vocal sac; hence the simple source assumption is well-justified. At the highest frequencies analyzed in this work, near 6 kHz, the wavelength is about 5.7 cm, or 2.85 times the width of the vocal sac, and the simple source assumption is less well-justified. The vocal sac can be larger, about 2.5 cm for calls with strong chucks,[41] which were not observed in this work. The effective volume of the vocal sac during the whine of a call

with no chuck is about 3000 mm^3 which yields an effective spherical radius of about 1 cm. Based on these dimensions and the simple source assumption, the frog's acoustic radiator was modeled as a uniformly pulsating sphere of radius 1 cm.

The túngara frog always calls in shallow water, often in small pools, puddles, or near the edge of ponds, with its vocal sac mostly above the surface of the water.[22] The acoustic ramification is that water, despite being fluid, appears as a nearly perfect rigid acoustic reflector to waves incident in air[38] and its surface is smooth and flat. Further, these small pools or pond edges provide a finite-sized, flat, rigid reflecting surface, bounded by soil and vegetation, which has acoustically soft surface properties. A wide variety of earth surfaces, from grassland, to cultivated earth, to layered forest floors exhibit similar acoustic properties, when subjected to transient incident acoustic pulses from above, with measured specific acoustic impedances that range from about eight times that of air at 1 kHz, to about two times that of air at 10 kHz.[44] The petri dish and table used in this work were intended to provide an idealized finite sized rigid reflecting plane, bounded by acoustically soft material, the air surrounding the table.

The models presented below demonstrate that the basic features of the observed radiation patterns are due to the frog behaving acoustically as a small pulsating sphere (the vocal sac) calling just above a finite-sized acoustically rigid plane (the water surface in nature or the table in this work), surrounded by an acoustically soft surface (soil and vegetation in nature or air in this work).

A. Simple source above an infinite rigid plane

It is also useful to demonstrate that the observed directivity is dependent on the size of the acoustically rigid reflecting surface being finite, so the first model shown is that of a simple source over an *infinite* rigid plane. Using the image method, the directivity of the field produced by this configuration is[45]

$$H(\theta) = 20 \log_{10}[\cos(kh \sin \theta)] \text{ (dB)}, \quad (2)$$

where $k=2\pi f/c$, the distance of the source above the plane is h, and θ is the angle above horizontal. The sound speed in air was 343 m/s. Setting $h=1$ cm, which corresponds to the 1 cm radius source described above sitting directly on the plane, and letting $f=2$ kHz, which is about the frequency of the second harmonic in this work, one finds very little directivity, as shown in Fig. 7(a). Increasing the frequency to $f=6$ kHz, also shown in Fig. 7(a), results in about 7 dB difference between 0° and 90°, but the amplitude is lower directly above the source, which is the opposite of that observed in Sec. III, where the radiated level was greater directly above the frog for the first, second, and third harmonics. This result indicates that the source over a rigid plane is not sufficient to explain the observed directivity presented in Sec. III.

B. Simple source above a finite rigid plane bounded by air

The general nature of the experimental apparatus used in Sec. III was simulated using a commercially available finite element software package. A two-dimensional axisymmetric finite element solution of the Helmholtz equation was obtained in a hemispherical domain. The coordinate axes of this domain, the radial dimension r and the height above the reflecting plane z, and the simulation geometry are schematized in Fig. 7(b). The center of a spherical 1 cm radius source with a uniform prescribed velocity was placed at $r=0$ and $z=1$ cm, as shown with the open circle (size exaggerated). The source was placed above a rigid circular surface that resided in the r-plane at $z=0$, with radius $r_0=15$ cm and thickness 5 cm extending below the r-plane. The table used in the measurements in Sec. III had the same thickness and its short side occupied the same radial dimension as shown, but was rectangular, whereas the table in the simulation is circular when rotated about the axis of symmetry (z-axis). This concession was made to allow efficient computation via a two-dimensional axisymmetric domain. A rectangular table would have required a computationally-intensive three-dimensional domain. The remaining domain was filled with air (sound speed of 343 m/s and density of 1.2 kg/m^3) and terminated at $r=1$ m with an outgoing spherical radiation condition. The simulation domain occupied $+90° < \theta < -90°$, although only the upper quadrant is shown in Fig. 7.

The simulation was run at several frequencies ranging from 1 to 6 kHz and the SPL was calculated at $r=1$ m for $0 < \theta < +90°$. This mimics the location of the microphones used in the directivity measurements in Sec. III. The resulting beam patterns are shown in Fig. 7(b). All curves were normalized to 0 dB at the angle of their maximum value. At 1 kHz, the radiation is directed above, is about 8 dB greater than along the horizontal, and is very similar to the measured radiation pattern shown in Fig. 3(b) and Figs. 5(a)–5(d) for the first harmonic, which was also about 1 kHz. As the frequency increases, both simulation and measurement show that radiation can be directed both above and at other angles, and that localized minima can form. Compare this to the relatively omnidirectional radiation seen in Fig. 7(a) at 2 kHz for the source above an infinite rigid plane and the lack of localized minima at either frequency. A finite-sized rigid reflecting plane is required to achieve both upward vertical directivity and localized minima.

C. Simple source above a finite rigid plane bounded by idealized earthen surfaces

The following models were undertaken as steps toward simulating a few aspects of the frog's natural calling environment. The finite element simulation described in Sec. IV B was repeated with the following changes: All calculations reported in this section were for a frequency of 2 kHz. The domain was reduced to $0 < \theta < +90°$. The flat, rigid reflecting surface below the source was retained, but instead of bounding it with air, the material properties and surface roughness of the natural environment were simulated. To be-

gin though, a flat rigid plane extending the entire length of the r-axis was used to serve as a comparison to the analytical solution for the simple source above an infinite rigid plane discussed in Sec. IV A. The simulation result, labeled "flat rigid" in Fig. 7(c), agrees very well with the analytical solution shown in Fig. 7(a). This validates the finite element model and indicates that the model source radiating above a flat rigid infinite plane produces nearly omnidirectional radiation at 2 kHz.

Next, the effect of surface roughness was investigated. The flat rigid plane was retained out to $r_0=15$ cm, but for 15 cm$<r_0<1$ m, the flat rigid surface was replaced with a random rough rigid surface. The location of this surface is indicated in Fig. 7(c) by the label "idealized earth." The rms surface roughness was 1.6 cm. The resulting directivity is shown in Fig. 7(c) by the curve labeled "rough rigid." The level is now about 8 dB less along the horizontal than directly above.

The material below the rough surface (15 cm$<r_0<1$ m) was then given acoustically soft material properties to mimic soil and vegetation. A specific acoustic impedance four times that of air, $4z_{\text{air}}$, was used (sound speed of 686 m/s and density of 2.4 kg/m^3) as is appropriate for a variety of soils at 2 kHz,[44] and the layer was extended to $z=-10$ cm, bounded on the bottom by a rigid boundary. The resulting directivity is shown in Fig. 7(c) with the curve labeled "rough soft." Now, the level directed upward is 20 dB higher than along the horizontal.

Finally, the rigid flat surface along the r-axis was replaced, and a 2-cm-thick, flat layer of the same acoustically soft material, with a specific acoustic impedance four times that of air ($4z_{\text{air}}$, sound speed of 686 m/s, and density of 2.4 kg/m^3), was placed on top of it for 15 cm$<r_0<1$ m. The resulting radiation pattern is shown in Fig. 7(c) by the curve labeled "flat soft." Again, one finds more radiation directed up than along the horizontal, by about 13 dB.

The effect of the soft layer's specific acoustic impedance was also investigated by varying it from eight times the specific acoustic impedance of air, to twice that of air, which is the range of surface acoustic properties found in Ref. 44. The shapes of the radiation patterns were very similar to the flat soft curve in Fig. 7(c), but with slightly different absolute values. For example, the differences between upward and horizontal radiation levels were 12.7, 13, and 11.7 dB, as the layer's impedance was varied from $2z_{\text{air}}$, to $4z_{\text{air}}$, to $8z_{\text{air}}$, respectively. The upward directivity is not strongly dependent on the surface properties of the material surrounding the reflecting surface (the water surface in nature), within the expected range of values for a variety of soils.

The effect of the size of the reflecting surface was also investigated. Its radius is r_0, as shown in Fig. 7(c). The model was run for 5 cm$<r_0<50$ cm, which corresponds to a range of $0.29<r_0/\lambda<2.9$ when normalized by the acoustic wavelength in air. The layer's acoustic properties were set at four times the specific acoustic impedance of air (sound speed of 686 m/s and density of 2.4 kg/m^3). The shape of the directivity curves changed as r_0 changed, but upward directivity was present in all cases. The difference in level between upward and horizontal radiation ranged between 8 and 16 dB. For $r_0<\lambda$, the radiation pattern was dipole-like, with the maximum level radiated directly upward and the level monotonically decreasing toward the horizontal. For $r_0>\lambda$, the radiation patterns developed local maxima and minima, or lobes, and the number of lobes increased as r_0 increased. The difference between local maxima and minima were typically about 4 dB. These results indicate that the size of the reflecting surface (the pool of water in nature) affects the specifics of the radiation patterns, but it does not affect the presence of upward directivity, for reflecting surfaces that are up to 50 cm in radius. Upward directivity would eventually be lost for increasing r_0, as was shown for the infinite reflecting plane in Figs. 7(a) and 7(c), but recall that túngara frogs call from shallow water,[22] which limits either the size of the puddle or the distance of the frog from the edge of a large pond.

The models shown in this section indicate that a frog calling just above a finite-sized acoustically rigid surface, such as the table in the present measurements or a shallow pool of water in nature, bounded by an acoustically soft material or by a rough surface, will result in more acoustic energy being directed upward than along the horizontal. Since the túngara always calls in shallow water, either near the edge of a pond or in a small pool, it is likely that upward directivity will be found in nature, as was found in the idealized environment used in Sec. III. Since the geometry of the natural pools, and the acoustic properties of the various soils and vegetation that surround the pools are not constant, many details of the túngara radiation patterns found in nature will differ from place to place, but upward-directed radiation patterns will likely persist. The details of these radiation patterns would also depend on nonuniform surface vibration of the vocal sac and acoustic interaction with the parts of the frog's body that were not modeled here (head, legs, and body), hence potentially explaining the variability among individuals already observed in Sec. III.

V. CONCLUSIONS

Mating calls of male túngara frogs were recorded in an anechoic chamber using an ultrasound-bandwidth microphone and using an audio-bandwidth microphone array oriented to observe acoustic directivity in the elevation angle (the vertical plane). The frogs produced calls, frequency-modulated whines, which were found to contain narrowband harmonics. No coherent signal was observed in the whines above 11.5 kHz. Thus, unlike the calls of some frogs,[32] the whines of túngara frogs studied here do not contain information in the ultrasonic range. These calls do exhibit substantial broadband and narrowband directivity. There was broadband directivity, expressed through the relative SPL of the entire whine. Directly above the frog, the radiated SPL was typically 6 dB greater than that radiated near the horizontal direction. Narrowband directivity was also seen in many of the harmonics of the calls. Higher-frequency harmonics displayed an increased directivity, with a 10 to 20 dB difference in radiated amplitude between angular directions in the vertical plane. Some of the harmonics were directed 45° from the ground, while other harmonics projected directly above

the calling frog. Finally, there were considerable differences observed from call-to-call, for a single frog at a single azimuthal angle. There were also differences seen as a function of azimuthal angle and certainly differences among individuals.

The models presented in Sec. IV indicate that the directivity observed in the idealized laboratory environment is due to the presence of a finite-sized, acoustically-hard, flat reflecting surface underneath the calling frog. This surface was created by the table used in the measurements, and is acoustically similar to the water surface from which the frogs call in nature. If this surface is bounded by an acoustically soft material or by a rough surface, both of which are found in the frog's natural environment, then acoustic radiation will be directed upward at levels higher than along the horizontal. This acoustic radiation pattern presents an evolutionary dilemma for the calling frog. A male's mating success is dependent on projecting the mating call into the active space for females, which is the horizontal plane. Yet due to the call directivity observed in the laboratory and predicted to exist in nature, the active space is greater in the vertical plane where the bats and flies reside. Assuming these radiation patterns are called amplitude independent, any increase in call amplitude would asymmetrically increase the caller's exposure to eavesdroppers compared to mates, causing a relative increase in mortality risk for the caller. The directivity pattern of the sound field is one component of the frog's communication system that is subject to the competing costs and benefits of communicating. Thus understanding the biophysics of the communication system is necessary for a deeper understating of both its function and evolution.

ACKNOWLEDGMENTS

We thank T. Hollon and K. Miller for their help with frog video analysis. We appreciate the comments of Michael Owren and one anonymous reviewer that greatly improved the manuscript. This study was funded by NSF Grant No. IBN-0078150 and The University of Texas at Austin Cockrell School of Engineering.

[1]C. E. Shannon, "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 379–423 (1948).
[2]W. W. L. Au, *The Sonar of Dolphins* (Springer-Verlag, New York, 1993).
[3]M. Hunter, Jr., A. Kacelnik, J. Roberts, and M. Vuillermoz, "Directionality of avian vocalizations: A laboratory study," Condor **88**, 371–375 (1986).
[4]A. Michelsen and P. Fonseca, "Spherical sound radiation patterns of singing grass cicadas, *Tympanistalna gastrica*," J. Comp. Physiol., A **186**, 163–168 (2000).
[5]K. Frommolt and A. Gebler, "Directionality of dog vocalizations," J. Acoust. Soc. Am. **116**, 561–565 (2004).
[6]H. Carl Gerhardt, "Sound pressure levels and radiation patterns of the vocalizations of some North American frogs and toads," J. Comp. Physiol., A **102**, 1–12 (1975).
[7]M. S. Dantzker, G. B. Deane, and J. W. Bradbury, "Directional acoustic radiation in the strut display of male sage grouse *Centrocercus urophasianus*," J. Exp. Biol. **202**, 2893–2909 (1999).
[8]M. Zuk and G. R. Kolluru, "Exploitation of sexual signals by predators and parasitoids," Q. Rev. Biol. **73**, 415–438 (1998).
[9]T. M. Peake, "Eavesdropping in communication networks," in *Animal Communication Networks*, edited by P. K. McGregor (Cambridge University Press, Cambridge, 2005), pp. 13–37.
[10]W. H. Cade, "Acoustically orienting parasites: Fly phonotaxis to cricket song," Science **190**, 1312–1313 (1975).
[11]S. Sakaluk and J. Belwood, "Gecko phonotaxis to cricket calling song: A case of satellite predation," Anim. Behav. **32**, 659–662 (1984).
[12]J. Bellwood and G. K. Morris, "Bat predation and its influence on calling behavior in neotropical katydids," Science **238**, 64–67 (1987).
[13]K. Schniederkötter and R. Lakes-Harlan, "Infection behavior of a parasitoid fly, *Emblemasoma auditrix*, and its host cicada *Okanagana rimosa*," J. Insect Sci. **4:36**, 1–7 (2004).
[14]G. R. Allen and J. Hunt, "Larval competition, adult fitness, and reproductive strategies in the acoustically orienting *Ormiine Homotrixa alleni* (Diptera: Tachinidae)," J. Insect Behav. **14**, 283–297 (2001).
[15]G. U. C. Lehmann, K.-G. Heller, and A. W. Lehmann, "Male bushcrickets favoured by parasitoid flies when acoustically more attractive for conspecific females (Orthoptera: Phanopteridae/Diptera: Tachinidae)," Entomol. Gen. **25**, 135–140 (2001).
[16]P. D. Bell, "Acoustic attraction of herons by crickets," J. New York Entomol. S. **87**, 126–127 (1979).
[17]M. D. Tuttle, L. K. Taft, and M. J. Ryan, "Acoustic location of calling frogs by *Philander opossums*," Biotropica **13**, 233–234 (1982).
[18]T. Halliday, *Sexual Strategy* (University of Chicago Press, Chicago, 1980).
[19]J. A. Endler, "Evolutionary implications of the interaction between animal signals and the environment," in *Animal Signals*, edited by Y. Espmark, T. Amudsen, and G. Rosenqvist (Tapir Academic, Trondheim, Norway, 2000).
[20]R. H. Wiley and D. G. Richards, "Adaptations for acoustic communication in birds: sound transmission and signal detection," in *Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Academic, New York, 1982), Vol. **1**.
[21]M. J. Ryan and N. M. Kime, "Selection on long-distance acoustic signals," in *Acoustic Communication*, edited by A. M. Simmons, A. N. Popper, and R. R. Fay (Springer, New York, 2003).
[22]M. J. Ryan, "The túngara frog," *A Study in Sexual Selection and Communication* (University of Chicago Press, Chicago, 1985).
[23]M. J. Ryan and A. S. Rand, "Mate recognition in túngara frogs: A review of some studies of brain, behavior, and evolution," Acta Zool. Sinica **49**, 713–726 (2003).
[24]M. Gridi-Papp, A. S. Rand, and M. J. Ryan, "Complex call production in túngara frogs," Nature (London) **441**, 38 (2006).
[25]M. D. Tuttle and M. J. Ryan, "Bat predation and the evolution of frog vocalizations in the Neotropics," Science **214**, 677–678 (1981).
[26]X. E. Bernal, A. S. Rand, and M. J. Ryan, "Acoustic preferences and localization performance of blood-sucking flies (*Corethrella* Coquillett)," Behav. Ecol. **17**, 709–715 (2006).
[27]X. E. Bernal, R. A. Page, A. S. Rand, and M. J. Ryan, "Cues for eavesdroppers: Do frog calls indicate prey density and quality?," Am. Nat. **169**, 412–415 (2007).
[28]K. S. Lynch, D. C. Crews, M. J. Ryan, and W. Wilczynski, "Hormonal state influences aspects of female mate choice in the túngara frog (*Physalaemus pustulosus*)," Horm. Behav. **49**, 450–457 (2006).
[29]M. J. Ryan and A. S. Rand, "Evoked vocal responses in male túngara frogs: Preexisting biases in male responses?," Anim. Behav. **56**, 1509–1516 (1998).
[30]O. Cramer, "The variation of the specific heat ratio and the speed of sound in air with temperature, pressure, humidity, and CO_2 concentration," J. Acoust. Soc. Am. **93**, 2510–2516 (1993).
[31]H. E. Bass, L. C. Sutherland, A. J. Zuckerwar, D. T. Blackstock, and D. M. Hester, "Atmospheric absorption of sound: Further developments," J. Acoust. Soc. Am. **97**, 680–683 (1995).
[32]A. Feng, P. Narins, C. Xu, W.-Y. Lin, Z.-L. Yu, Q. Qiu, Z.-M. Xu, and J.-X. Shen, "Ultrasonic communication in frogs," Nature (London) **440**, 333–336 (2006).
[33]D. Aylor, "Noise reduction by vegetation and ground," J. Acoust. Soc. Am. **51**, 197–205 (1972).
[34]W. T. Fitch, J. Neubauer, and H. Herzel, "Calls out of chaos: the adaptive significance of nonlinear phenomena in mammalian vocal production," Anim. Behav. **63**, 407–418 (2002).
[35]A. B. Pippard, "The driven anharmonic vibrator; subharmonics; stability," *The Physics of Vibration* (Cambridge University Press, Cambridge, 1978), Vol. **1**, Chap. 9.
[36]M. B. Hindmarsh and D. J. Jefferies, "On the motions of the offset impact oscillator," J. Phys. A **17**, 1791–1804 (1984).
[37]V. Tournat, V. E. Gusev, and B. Castagnède, "Subharmonics and noise excitation in transmission of acoustic wave through unconsolidated granular medium," Phys. Lett. A **326**, 340–348 (2004).
[38]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals*

of Acoustics, 3rd ed. (Wiley, New York, 1982).
[39]W. A. Watkins, E. R. Baylor, and A. T. Bowen, "The call of eleutherodactylus johnstonei, the whistling frog of bermuda," Copeia **1970**, 558–561 (1970).
[40]W. E. Duellman and L. Trueb, *Biology of Amphibians* (McGraw-Hill, New York, 1986).
[41]R. Dudley and A. S. Rand, "Sound production and vocal sac inflation in the túngara frog, Physalaemus pustulosus (Leptodactylidae)," Copeia **1991**, 460–470 (1991).
[42]A. S. Rand and R. Dudley, "Frogs in helium: The anuran vocal sac is not a cavity resonator," Physiol. Zool. **66**, 793–806 (1993).
[43]K. D. Wells, *The Ecology & Behavior of Amphibians* (The University of Chicago Press, Chicago, 2007).
[44]C. G. Don and A. J. Cramond, "Soil impedance measurements by an acoustic pulse technique," J. Acoust. Soc. Am. **77**, 1601–1609 (1985).
[45]D. T. Blackstock, *Fundamentals of Physical Acoustics* (Wiley, New York, 2000).

Vocalizations of wild Asian elephants (*Elephas maximus*): Structural classification and social context

Smita Nair and Rohini Balakrishnan[a)]
Centre for Ecological Sciences, Indian Institute of Science, Bangalore, Karnataka 560012, India

Chandra Sekhar Seelamantula
Department of Electrical Engineering, Indian Institute of Science, Bangalore, Karnataka 560 012, India,

R. Sukumar
Centre for Ecological Sciences, Indian Institute of Science, Bangalore, Karnataka 560012, India

(Received 26 November 2008; revised 14 August 2009; accepted 15 August 2009)

Elephants use vocalizations for both long and short distance communication. Whereas the acoustic repertoire of the African elephant (*Loxodonta africana*) has been extensively studied in its savannah habitat, very little is known about the structure and social context of the vocalizations of the Asian elephant (*Elephas maximus*), which is mostly found in forests. In this study, the vocal repertoire of wild Asian elephants in southern India was examined. The calls could be classified into four mutually exclusive categories, namely, trumpets, chirps, roars, and rumbles, based on quantitative analyses of their spectral and temporal features. One of the call types, the rumble, exhibited high structural diversity, particularly in the direction and extent of frequency modulation of calls. Juveniles produced three of the four call types, including trumpets, roars, and rumbles, in the context of play and distress. Adults produced trumpets and roars in the context of disturbance, aggression, and play. Chirps were typically produced in situations of confusion and alarm. Rumbles were used for contact calling within and among herds, by matriarchs to assemble the herd, in close-range social interactions, and during disturbance and aggression. Spectral and temporal features of the four call types were similar between Asian and African elephants.
 [DOI: 10.1121/1.3224717]

PACS number(s): 43.80.Ka, 43.66.Gf [MJO] Pages: 2768–2778

I. INTRODUCTION

A number of mammalian groups, including carnivores, ungulates, cetaceans, rodents, and primates, use acoustic signals for intraspecific communication in a variety of contexts such as mate-finding, courtship, aggression, alarm, and territoriality (Bradbury and Vehrencamp, 1998). Elephants, which are the largest terrestrial mammals, also use acoustic signals to communicate over both short-medium (Douglas-Hamilton, 1972; Krishnan, 1972; McKay, 1973; Poole and Moss, 1989; Poole *et al.*, 1988; McComb *et al.*, 2003) and long distances (Garstang *et al.*, 1995; Larom *et al.*, 1997). Whereas acoustic signaling in the African elephant (*Loxodonta africana*) has been extensively studied in their savannah habitats (Poole and Moss, 1989; Payne, 1998; Poole *et al.*, 1988; Poole, 1999; Langbauer, 2000), very little is known about acoustic signaling in Asian elephants (*Elephas maximus*), which are largely forest dwelling. In this paper, the vocalizations of wild Asian elephants from southern India are presented.

The Asian and African elephants are believed to have diverged from a common ancestor in the African continent about $(5–6)\times10^6$ yrs. ago (Maglio, 1973; Sukumar, 2003). Comparing the vocal repertoires and social contexts of the calls of the two species could give us insight into the pattern of evolution of signal structure and function in an acoustic communication system. Elephants, both African and Asian, live in groups and have a complex social structure (Moss and Poole, 1983; Sukumar, 1989). Elephant herds typically consist of related females and their offspring, including subadult males (Moss, 1988; Vidya and Sukumar, 2005). Adult males of both Asian and African elephants are often solitary and wander widely, joining herds temporarily in the context of reproduction. Males are also known to temporarily associate with other males to form all-male groups (Poole, 1994; Sukumar 1989). A higher level of social organization, with herds organized into clans, has been observed in both Asian and African elephant populations (Moss and Poole, 1983; Sukumar, 1989, 2003; Wittemyer *et al.*, 2005). Given that elephants may move large distances in search of food and water, acoustic signaling serves as an effective way of communication between males and females, among herd members, and between herds (Poole *et al.*, 1988).

Several studies have characterized the vocal repertoire of both captive and wild African elephants both in terms of the structure of calls and social context (Berg, 1983; Poole *et al.*, 1988; Langbauer, 2000; Leong *et al.*, 2003; Wood *et al.*, 2005; Soltis *et al.*, 2005, Stoeger-Horwath *et al.*, 2007; Leighty *et al.*, 2008) and their transmission in space (Garstang *et al.*, 1995; Larom *et al.*, 1997). The functions of some of the calls have also been examined using playback experiments in the field (Langbauer *et al.*, 1991; Poole,

a)Author to whom correspondence should be addressed. Electronic mail: rohini@ces.iisc.ernet.in

1999; McComb *et al.*, 2000, 2001, 2003). Acoustic signatures for individual recognition and the capacity of elephants to differentiate between the calls of different individuals have also been studied, largely in the context of low-frequency rumbles (McComb *et al.*, 2000, 2001; Clemins *et al.*, 2005; Soltis *et al.*, 2005). In contrast, very little is known about either the vocal repertoire or the social contexts and functions of vocalizations in Asian elephants, even though the low-frequency rumbles with their infrasonic components were in fact first recorded in captive Asian elephants (Payne *et al.*, 1986). So far only a brief description of the calls of wild Asian elephants is available (McKay, 1973).

This study presents the first recordings of the vocalizations of free-ranging, wild Asian elephants and classifies them into call types based on quantification of the structural characteristics of the calls. The social contexts of these call types, as well as the structural diversity within one of these call types—the rumble—were examined. This study is meant to provide a baseline description of Asian elephant vocalizations for more detailed field studies along the lines of what has been achieved for the African elephant.

II. METHODS

A. Study area

The study was carried out in the Mudumalai Wildlife Sanctuary (321 km^2, 11°33′ to 11°39′N and 76°23′ to 76°43′E) located in the Western Ghats of southern India. The major types of vegetation found here are tropical semi-evergreen, moist deciduous, dry deciduous, and dry thorn forests corresponding to a distinct rainfall gradient of higher rainfall (~1800 mm annually) in the southwest to lower rainfall (~600 mm annually) in the northeast. Mudumalai, with an average elephant density of about 2 individuals/km^2 (Varman and Sukumar, 1995, and unpublished results for 1993–2006) is part of the Nilgiri Biosphere Reserve as well as Project Elephant Range No. 7 that holds the single largest population of the Asian elephant globally (Sukumar, 2003; Venkataraman *et al.*, 2002).

B. Behavioral and acoustic data collection

Fieldwork was conducted mainly in the dry months (February to May) of 2006 and 2007, in and around waterholes, salt licks, open areas such as swamps or grasslands, and within the forests. Surveys were carried out by vehicle and on foot to locate elephant herds during the day. There were a total of 214 encounters with herds during these surveys. The herds were classified into three major categories, namely, mixed herds (both females and males in all age classes), female herds (no adult or sub-adult males) and male herds (adult and sub-adult males only). Solitary individuals, male or female, were also occasionally encountered. The proportion of adult (>15 years) and sub-adult (5–15 years) males in this population is estimated to be 12.5% (Arivazhagan and Sukumar, 2005) and approximately 11% from our encounters. Elephants were individually identified based on distinguishing characteristics such as ear-fold, tail characteristics, and pigmentation. The age class of individuals was also determined using methods described elsewhere (Sukumar, 1989). Adult and sub-adult individuals were grouped together in a single category (Adults) and juveniles and calves were grouped in a single category (Young).

Acoustic data were acquired using a TASCAM DA-P1 Digital Audio Tape (DAT) recorder (frequency response of 1 Hz–20 kHz) and AKG CK 62-ULS omni-directional condenser microphone–C480B preamplifier combination (frequency response of 10 Hz–20 kHz) at a sampling frequency of 48 kHz. The distance to the herd from the observer ranged from 5 m to 75 m. Sources of noise like wind and vehicle were noted for filtering during data analysis. In addition, video recordings were made using a Canon MV1 Digital Video Camcorder in relatively open areas. The date, time, and location of recordings were recorded. Detailed behavioral observations were carried out along with the audio recordings to categorize elephant behavior for association with different call types. Vocalizing individuals were identified on the basis of typical postures and behaviors (www.elephantvoices.org), for example, position of the trunk and body movements (Table IV), exhibited during calling. The presence of other species in the vicinity of elephant herds was also noted.

C. Acoustic data analysis

The recorded signals were re-digitized using a Creative Sound Blaster A/D card (16 bit) at a sampling rate of 48 kHz and the calls were analyzed for their spectral and temporal properties using MATLAB version 6.5 (The Mathworks, Inc., Natick, MA) and PRAAT 5.1.07 (www.praat.org, Paul Boersma and David Weenink).

1. Estimation of call duration

Noise was removed using the spectral subtraction method (Boll, 1979). A custom-written program was used to automatically detect the onsets and cessations of calls. The high frequencies in the signal beyond one-tenth of its bandwidth were suppressed using a low-pass filter (1024-point, finite impulse response). To compute for the temporal envelope, an 85 ms Hamming window normalized to unity sum was used for smoothing. The modulus of the resulting low-pass signal was then convolved with the smoothing window to obtain an approximate envelope, which has large values in the regions corresponding to the signal. The temporal envelope was further subjected to time-domain thresholding with a hard-threshold criterion, which clamps the values of the envelope smaller than 20% of the peak magnitude to zero. The onsets and cessations of calls could be determined from the thresholded envelope.

2. Estimation of minimum and maximum frequencies

A narrow-band spectrogram was computed and the minimum and maximum frequencies of the calls were determined using their power spectral densities. To estimate the essential bandwidth, the power spectral density (PSD) of the noisy signal was first computed using a Hamming window (20 ms). The PSD of the noise was then estimated using the initial 1 s of the recording. The estimated PSD of the noise was subtracted from that of the noisy signal to yield an esti-

TABLE I. List of measured call features.

Acoustic measures	Definition
Mean F_0	Mean frequency of the fundamental measured over the duration of the call (Hz)
Maximum F_0	Maximum frequency of the fundamental (Hz)
Minimum F_0	Minimum frequency of the fundamental (Hz)
Maximum call frequency	Highest frequency of the call (Hz)
Minimum call frequency	Lowest frequency of the call (Hz)
Frequency range of call	Maximum–minimum call frequency
Duration	Length of call in seconds
Harmonicity	Harmonics-to-noise ratio of call
Peaks 1–7 frequency	Frequencies of the first through seventh spectral peaks in the LPC spectrum
Peaks 1–7 amplitude	Amplitudes of the first through seventh spectral peaks in the LPC spectrum
Rumble start frequency	Frequency of the fundamental at the start of the call (Hz)
Rumble end frequency	Frequency of the fundamental at the end of the call (Hz)
Rumble peaks	Number of observed peaks with a minimum frequency modulation of 4 Hz
Rumble percent to maximum	Time in percentage from signal onset to maximum frequency of fundamental
Rumble percent to minimum	Time in percentage from signal onset to minimum frequency of fundamental
Rumble mean modulation	Frequency change per unit time for the fourth harmonic

mate of the PSD of the signal. In general, the estimate had positive values at all discrete Fourier transform (DFT) bins; any negative values (due to estimation errors) were clamped to zero. From this estimate, the bandpass region containing about 80% of the total energy was considered as the frequency range of the signal.

For rumbles, however, calls were first low-pass-filtered with a cutoff of 250 Hz (this value was based on a preliminary inspection of the spectrograms). The signals thus obtained were re-sampled (by the zero padding approach) to obtain more points along the frequency axis (Oppenheim and Schafer, 1989). Spectrograms of these signals (DFT with Hamming window size of 5 ms, 50% overlap) were used for further analysis.

3. Pitch, spectral envelope, and harmonicity analyses

Pitch, spectral envelope, and harmonicity analyses were carried out using PRAAT 5.1.07 (www.praat.org, Paul Boersma and David Weenink). Pitch analysis was carried out using a pitch floor of 10 Hz for rumbles and 100 Hz for most other calls. Spectral envelope analysis was carried out using linear predictive coding (LPC). Calls other than rumbles were re-sampled at 16 000 Hz whereas rumbles were re-sampled at 600 Hz. LPC analysis was carried out using the autocorrelation method with a time step of 0.005 s and a window length of 0.005 s (0.05 s for rumbles). Harmonicity (harmonics-to-noise ratio) was analyzed using the cross-correlation algorithm with a time step of 0.05 s except for chirps, for which 0.005 s was used due to the short duration of the calls.

4. Measured call features

Calls of different age-sex classes (adult male, adult female, young male, and young female) were analyzed separately. The measured features include the call frequency range, duration, harmonicity, as well as mean, minimum, and maximum values of the fundamental frequency (F_0) (Table I). In addition, the frequencies and amplitudes of the LPC peaks were measured to examine spectral patterning. The analyzed acoustic features were separated according to the four age-sex classes and compared statistically (for all categories where the sample size was ≥ 3) for each call type using a combination of one-way analysis of variance and pair-wise comparisons of means by unpaired *t*-tests. The fea-

TABLE II. Sample broken down by age-sex class and behavioral context.

			Number of female individuals		Number of male individuals			Number of calls	
Call type	Total number of calls	Total number of individuals	Adult	Young	Adult	Young	Individuals of unknown sex	Known context	Unknown context
Trumpet	77	37	27(58)	4 (9)	3 (5)	1 (1)	2 (4)	71	6
Roar	56	21	12 (42)	4 (6)	3 (3)	0	2 (5)	51	5
Chirp	68	25	18 (53)	⋯	4 (10)	⋯	3 (5)	66	2
Rumble	57	26	14 (27)	1(2)	0	2 (4)	9 (24)	56	1

Values in brackets are number of calls.

TABLE III. Spectral and temporal features of the four major call types.

Acoustic measures	Sex	Age group	Call type			
			Trumpet	Roar	Chirp	Rumble
Mean F_0 (Hz)	Female	Adult	677.5 ± 29.1	592.8 ± 93.7	Unmeasurable[a]	18.9 ± 1
		Young	787.8 ± 49.3	662.9 ± 165.2	Unknown[b]	20.5
	Male	Adult	828.1 ± 34.8	604.1 ± 53.7	Unmeasurable[a]	Unknown[b]
		Young	607.8	Unknown[b]	Unknown[b]	15.6
Minimum F_0 (Hz)	Female	Adult	607.4 ± 24.5	403.9 ± 51.3	Unmeasurable[a]	11.5 ± 0.7
		Young	697.3 ± 62	516.5 ± 130.4	Unknown[b]	11.8
	Male	Adult	745.6 ± 40.2	386.4 ± 99.1	Unmeasurable[a]	Unknown[b]
		Young	530.0	Unknown[b]	Unknown[b]	10.5
Maximum F_0 (Hz)	Female	Adult	864.60 ± 83.5	1214.5 ± 221.2	Unmeasurable[a]	29.4 ± 3.4
		Young	853.3 ± 56.4	751.2 ± 186.4	Unknown[b]	34.3
	Male	Adult	878.9 ± 47.5	1274.4 ± 471.6	Unmeasurable[a]	Unknown[b]
		Young	645.4	Unknown[b]	Unknown[b]	24.3
Call duration (s)	Female	Adult	0.7 ± 0.1	2 ± 0.3	0.2 ± 0.04	5.4 ± 0.6
		Young	0.7 ± 0.1	1.2 ± 0.3	Unknown[b]	2
	Male	Adult	1.3 ± 0.5	1.5 ± 0.4	0.2 ± 0.03	Unknown[b]
		Young	0.7	Unknown[b]	Unknown[b]	4.5
Harmonicity[c] (lower and upper quartile)	Female	Adult	6.7(5.3,8)	1.9(0.5,3.6)	−0.1(−1.2,4.7)	6.3(5.1,8.1)
		Young	6.3(5.6,7.1)	0.4(0.1,1.6)	Unknown[b]	4.8
	Male	Adult	6.8(6.6,7.6)	1.1(−0.2,2)	5.2(3.5,7.5)	Unknown[b]
		Young	2.1	Unknown[b]	Unknown[b]	5.9

Values are ± standard error. One call was randomly selected for individuals with multiple calls.
[a]Unable to extract fundamental.
[b]Call type not present in sample.
[c]Median values are reported for harmonicity since the distributions were skewed.

tures compared included mean F_0, minimum F_0, maximum F_0, call duration, harmonicity, and frequency and amplitude of the peaks of the LPC spectrum.

For the rumbles, the start and end frequencies of F_0, and percentage time from the start to the maximum and minimum frequencies were measured (Table I). To characterize the frequency modulation in finer detail two measures were used (Table I). The fourth harmonic of all calls was used since visual inspection of the spectrograms revealed that the modulation could be measured at high resolution across almost all recordings whereas the fundamental was sometimes contaminated with noise. The higher harmonics may also have greater functional relevance in social recognition as argued by McComb *et al.* (2003) based on playback experiments on wild African elephants. The first measure was the number of observed peaks in the fourth harmonic of each call, determined visually from the spectrogram with a minimum frequency modulation of 4 Hz as a cutoff. The second measure was the frequency change per unit time for the fourth harmonic, measured by the cumulative change in frequency divided by the time of the call (window length of 200 ms).

To detect the presence of structural subtypes within the rumbles, 13 measured call features (except harmonicity and LPC peaks) were used to generate pair-wise Euclidean distances between calls. The distance matrix was then subjected to cluster analysis using the Unweighted Pair Group Method with Arithmetic mean (UPGMA) (Sneath and Sokal, 1973; Manly, 1986). All statistical analyses were performed using STATISTICA (1999, Statsoft Inc., Tulsa, USA).

III. RESULTS

A total of 371 calls were recorded from 154 individuals. Of these, 258 calls were analyzed from 109 individuals, consisting of 14 males and 95 females (Table II). The remaining calls were discarded on account of low signal-to-noise ratio or overlap with other calls due to simultaneously vocalizing individuals. Based on structural characteristics, the calls could be classified into four types, namely, trumpets, chirps, roars, and rumbles (Fig. 1). One of the call types, the rumble, could be distinguished by its unique frequency range (10–173 Hz, Fig. 1, G–I), which was much lower than that of the other calls. Chirps were distinguished by their unique temporal structure (Table III, Fig. 1, J–K): they were typically of much shorter duration (0.2 s ± 0.1 s) than the other calls. Trumpets and roars differed from chirps in their duration (mean=1 s and 2 s, Table III). Although both trumpets and roars shared the same frequency range, as revealed by their spectral peaks (Fig. 2), the decrease in power with increasing frequency was steeper in roars. Roars also had significantly lower harmonicity than trumpets (Table III, Mann–Whitney U test, $U=313$, $Z=-7.8$, and $P<0.0001$).

There were no significant differences between the different age-sex classes in all of the call features that were examined in roars, chirps, and rumbles. The only significant differences were between adult male and female trumpets, wherein females had significantly lower mean F_0 and minimum F_0 (unpaired t-tests, $P=0.015$, $t=2.45$ and $P=0.043$, $t=2.78$) than males (Table III).

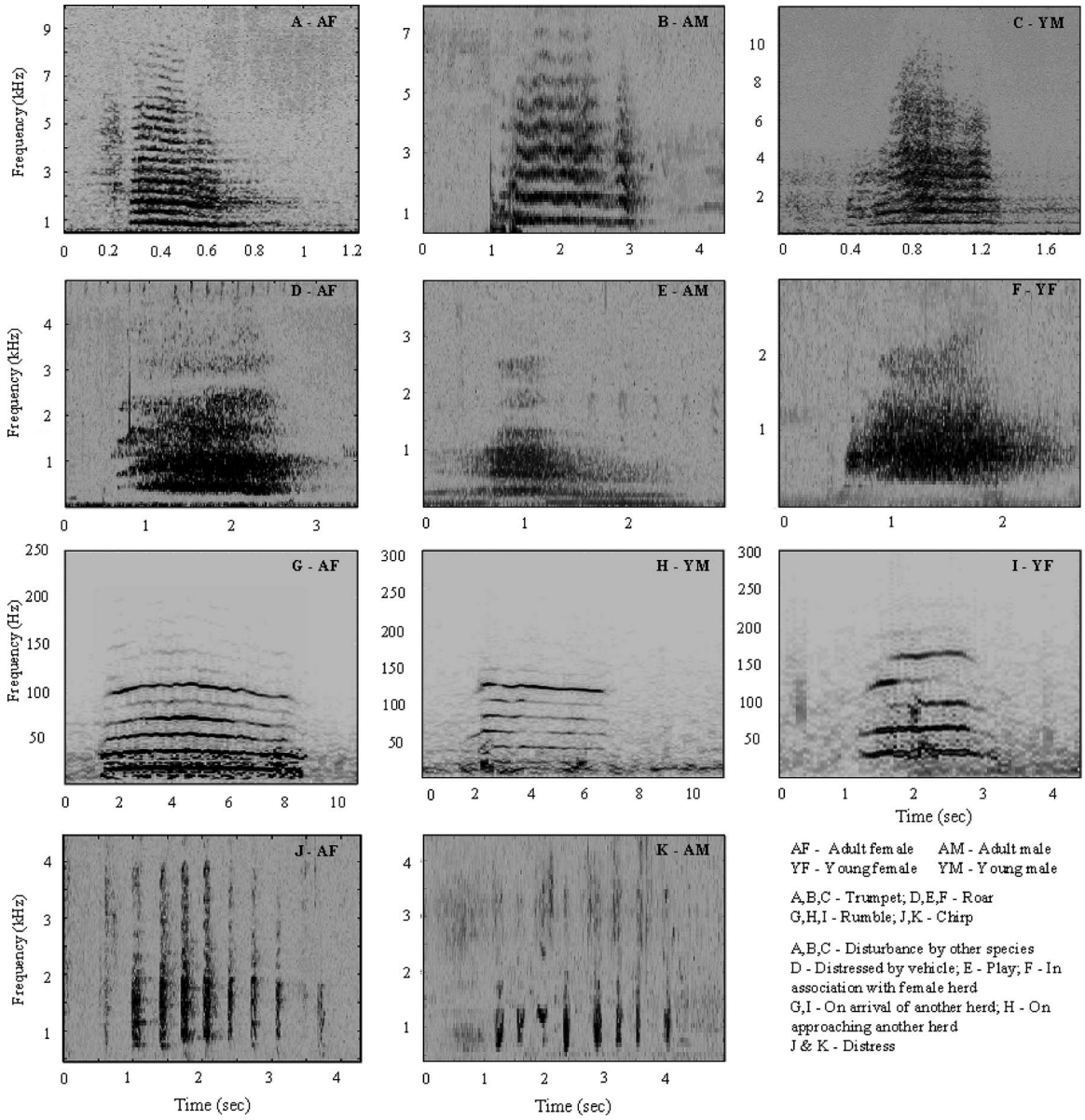


FIG. 1. Spectrograms of examples of the four call types. Sampling frequency for all calls was 48 kHz with 50% overlap. Window size is 100 ms for trumpet, roar, and chirp, and 200 ms for rumbles. Note the different *Y*-axis scale in G–I.

The different call types were associated with characteristic body postures and behaviors of the vocalizing individual and other members of the herd. These are described in detail in Table IV.

A. Trumpets

Trumpets were loud, conspicuous high-frequency calls. They were in the frequency range of 405–5879 Hz with a mean duration of about 1 s (Table V). They had a rich harmonic structure with at least seven clearly visible harmonics (Fig. 1, A–C). Spectral envelope analysis revealed the first frequency peak to be at 706 Hz (Fig. 2). The fourth frequency peak (at 3078 Hz) was about 13 dB lower in amplitude than the first frequency.

Out of 73 trumpets where the age-sex class was unambiguous, 58 (79.5%) were produced by adult or sub-adult females, nine (12.3%) by juvenile females, five (6.8%) by adult or sub-adult males, and one (1.4%) by a juvenile male (Table II). Out of the 71 trumpets where the context was clear, seven (9.9%) were in the context of play, 17 (23.9%) in the context of disturbance by humans or vehicles, 29 (40.8%) in the context of disturbance by other non-human species, ten (14%) in the context of inter-specific aggression, and eight (11.2%) while running out of a waterhole (Fig. 3). Specifically, trumpeting was observed while encountering other species such as deer (*Axis axis*), gaur (*Bos gaurus*), dhole (*Cuon alpinus*), bears (*Melursus ursinus*), tigers (*Panthera tigris*), egrets (*Egretta garzetta*), and humans.

B. Roars

Roars were noisy, long calls that were in the frequency range of 305–6150 Hz and had a mean duration of about 2 s (Table V, Fig. 1, D–F). They were in the same frequency range as trumpets and their spectral patterning was also simi-

TABLE IV. Social contexts of different call types.

Call type	Context	Vocalizing individual associated posture/behavior	Other herd members associated posture/behavior	Total number of calls
Trumpet	Play	Run after, lash out with trunk	Panic running, vocalize	7
		Head wagging, flop trunk on head	Bouts of agitated movement	
		kick back, test mouth	No visible response	
	Disturbed by humans/vehicle	Advance, J sniff	Bunch, panic running	17
		Foot stomping, redirected aggression	Vocalize, foot stomping, ear flapping	
		Charge, run away, panic running	Bouts of agitated movement	
		Bouts of agitated movement	Trunk twining with other individuals	
			Test mouth/temporal area of others	
	Disturbed by other species	Face other individuals, J sniff	Vocalize, bouts of agitated movement	29
		Run away, charge, panic running	Run away, ear flapping, stand still	
		Run after/raised trunk/trunk lashing	Bunch and advance	
		Bouts of agitated movement, redirected aggression	Huddle calf in between	
		Head shaking, ear flapping, foot stomp, and kick dust		
	Aggression	Charge, run away, lash out with trunk	Run away, retreat from, panic running	10
		Run after, J sniff	Vocalize, no visible response	
	Run out of water hole/exiting	Run out	Run out, no visible response	8
	Unknown context	Facing the landscape, panic running, run away	Vocalize, bouts of agitated movement	6
			Panic running, run away, no visible response	
Chirp	Disturbed by humans/vehicle	Head shaking, tail raised	Vocalize, bunch	30
	Disturbed by other species	Run in circles, redirected aggression	Panic running	26
	Separation from herd	Foot stomping, kick dust	No visible response	4
	Aggression within group	Run in circles, redirected aggression		6
	Unknown context			2
Roar	Play	Wallow, run after, lash out with trunk	Vocalize, wallow, panic running	7
	Disturbed by humans/vehicle	Advance toward, charge, panic running	Vocalize, bunch retreat from	4
	Disturbed by other species	Head shaking, ear flapping, foot stomping	Vocalize, bouts of agitated movement	28
		Charge, run away, redirected aggression	Run away, ear flapping, stand still	
		Panic running, bouts of agitated movement	Bunch and advance, bunch calf in between	
	Aggression	Pushing, dueling, run after, kick dust	Pushing, dueling	3
		Charge, run after	Retreat from, panic running	
	Facing another group/landscape	Stand still, face approaching herd or heterospecific	Vocalize, advance toward	9
		Advance toward, run after	No visible response	
	Unknown context	Panic running	Panic running	5
Rumble	Disturbed by humans/vehicle	Stand still, head shaking, foot stomping	Vocalize, bouts of agitated movement	20
		Bouts of agitated movement, run away	Grouping and exit, ear flapping	
		Scanning, ear spreading, redirected aggression	Group approaches	
		J sniff, ear flapping	Bunch around vocalizing individual	
		Face other individuals and vocalize	Huddle calf in between	
	Disturbed by other species	Stand still, bouts of agitated movement	Vocalize, J sniff, group and exit	6
		J sniff, redirected aggression	Ear flapping, head shaking, stand still	
		Head shaking, foot stomping, ear flapping	Panic running, run after	
	Let's go	Let's go stance	Group and exit area	3
	Contact call	Scanning listening, move ahead, vocalize	Vocalize, stand still, listening	3
	Interactive rumbling (intra- and inter-group interactions including greeting)	Stand still, scanning, pushing	Vocalize, group and advance toward	24
		Intermittent touching, trunk twining with conspecifics	Stand still, ear flapping	
		Testing mouth/temporal area of others	Advance toward, retreat from, follow	
		Ears spread, facing another group	Intermittent touching	
		Stand still followed by swing trunk movement	Testing mouth/temporal area of others	
	Unknown context	Running	Follow	1

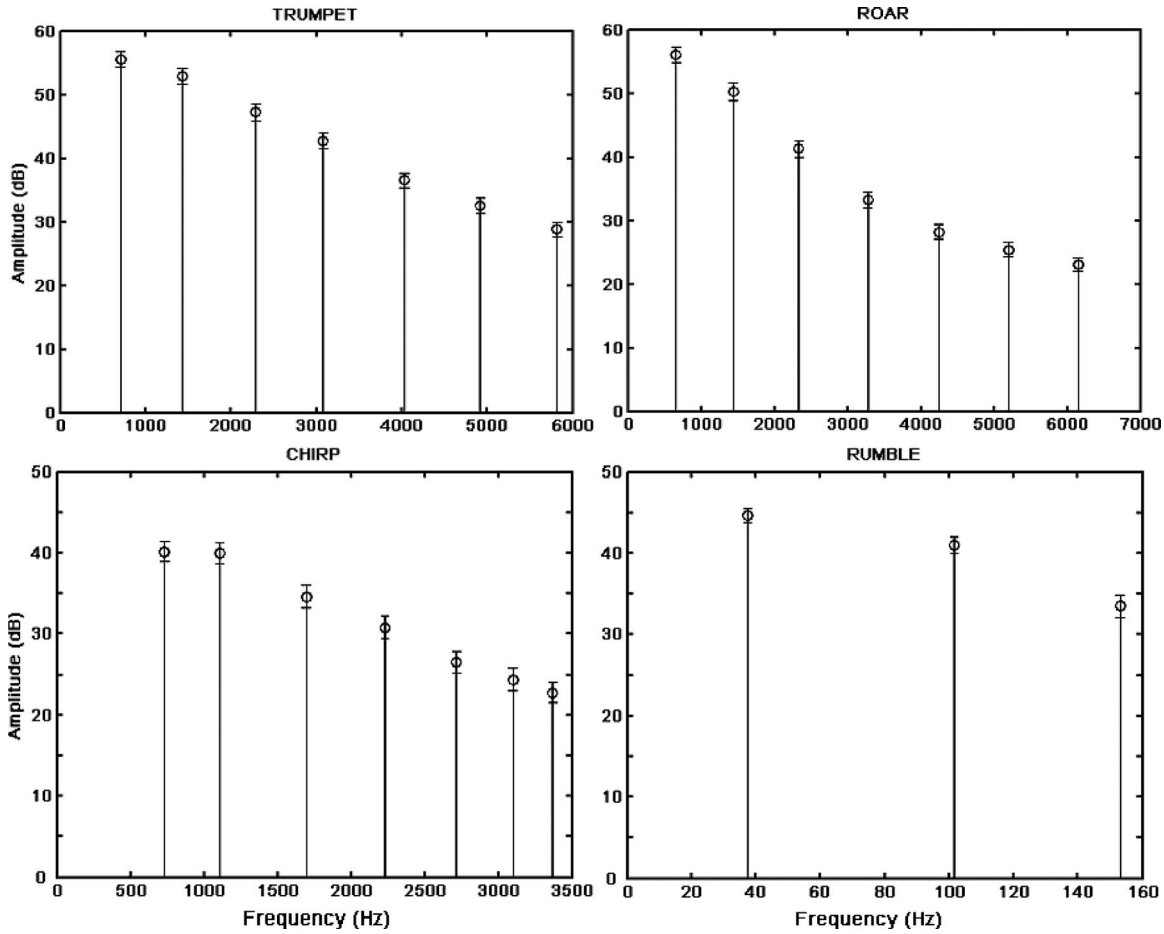


FIG. 2. Spectral envelopes of the four call types obtained using linear predictive coding. Values were pooled across the four age-sex classes since there were no statistically significant differences in the frequencies and amplitudes of peaks of the LPC spectrum in any of the call types.

lar with seven frequency peaks (Fig. 2), the first frequency peak being at 656 Hz. The amplitude fell steeply with increasing frequency, the fourth frequency peak being 23 dB below the first (Fig. 2). This was in contrast to trumpets, where the fourth frequency peak was about 13 dB below the first (Fig. 2). Roars also showed significantly lower harmonicity than trumpets (Table III). Both of the above are probably responsible for the unique perceptual quality of roars compared to trumpets.

Out of 51 calls where the age-sex class was clear, 42 (82.3%) were produced by adult or sub-adult females, six (11.7%) by juvenile females, and three (5.8%) by adult or sub-adult males (Table II). Out of 51 cases where the context was clear, seven (13.7%) were produced during play, four (7.8%) due to disturbance by humans or vehicles, 28 (54.9%) in the context of encounters with other non-human species, three (5.9%) during inter-specific aggressive interactions, and nine (17.6%) while facing another group or on entering a landscape (Fig. 3).

C. Chirps

Chirps were found to lie in the frequency range of 313–3370 Hz (Table V) and were produced in a series (Fig. 1, J–K) ranging from 2 to 8 (mean number=5.2 ± 2.6, $n=25$ individuals) in a single bout. The duration of a bout of chirping ranged from 0.68 s to 3.8 s. Spectral envelope analysis revealed up to seven discernible frequency peaks (Fig. 2), with the first two peaks having equal amplitude. Although chirps had seven frequency peaks, the range over which these peaks were distributed was much narrower (Fig. 2) than in the case of trumpets and roars. Chirps showed significantly lower harmonicity than trumpets and rumbles (Table III, Mann–Whitney U-test, $U=995.5$, $Z=-4.38$, $P<0.0001$, and $U=786.5$, $Z=-3.8$, $P<0.0001$).

Out of 63 chirp bouts where the age-sex class was clear, 53 (84%) were produced by adult or sub-adult females and ten (15.9%) by adult or sub-adult males (Table II). Out of 66 cases where the context was clear, 30 (45.5%) were due to disturbance by humans or vehicles, 26 (39.3%) due to disturbance by other non-human species, four (6%) in the context of separation of an individual from the herd, and six (9%) during intra-group aggression produced by an individual other than those directly involved in the aggression (Fig. 3).

D. Rumbles

Rumbles were the only call type in the repertoire with infrasonic components. Rumbles were found to lie in the frequency range of 10–173 Hz, with a mean duration of 5.2 s (Table V). They had a distinct harmonic structure (Fig. 1, G–I, Table III), similar to trumpets ($U=1508$, $Z=0.616$, and $P=0.54$). Spectral envelope analysis revealed three peaks,

TABLE V. Comparison of call features of Asian and African elephants

Call type	Call feature	Asian elephant	African elephant
Trumpet	Mean F_0 (Hz)	696.4 ± 20.3	390[a], 300[b]
	Frequency range call (minimum to maximum) (Hz)	405–5879	300–3 000[b]
	Dominant frequency[c]	706.3 ± 21.2	695[a]
	Duration (s)	0.9 ± 0.1	2[a], 1–5[b]
Roar	Mean F_0 (Hz)	649.5 ± 30.8	Unmeasurable[a]
	Frequency range call (minimum to maximum) (Hz)	305–6150	Unknown
	Dominant frequency[c]	656.5 ± 29	574[a]
	Duration (s)	2.0 ± 0.2	3.8[a]
Chirp (bark)	Mean F_0 (Hz)	Unmeasurable	Unmeasurable[a]
	Frequency range call (minimum to maximum) (Hz)	313–3370	Unknown
	Dominant frequency[c]	731.9 ± 26.9	629[a]
	Duration (s)	0.2 ± 0.01	0.47[a]
Rumble	Mean F_0 (Hz)	20.3 ± 0.7 (14–24[d])	27.7[e], 12[b]
	Frequency range call (minimum to maximum) (Hz)	10–173	12–200[b]
	Dominant frequency[c]	37.4 ± 2.9	Unknown
	Duration (s)	5.2 ± 0.3	4.1[e], 1–10[b]
Let's go	Mean F_0 (Hz)	18.5 ± 0.1	15[f]
Contact call	Mean F_0 (Hz)	21.2 ± 1.9	18[f]

Values are ± standard error.
[a]Berg (1983).
[b]Leong *et al.* (2003).
[c]The frequency peak with the highest amplitude in the LPC spectrum.
[d]Payne *et al.* (1986).
[e]Wood *et al.* (2005).
[f]Poole *et al.* (1988).

with the first peak at 37 Hz (Fig. 2). The power of the third peak was about 11 dB lower than that of the first frequency (Fig. 2).

Of the 33 cases where the identity of the rumbling individual was unambiguous, 27 (81.8%) were produced by adult or sub-adult females, two (6.1%) by juvenile females, and four (12.1%) by juvenile males (Table II). Out of 56 cases where the context of rumbling was clear, 20 (35.7%) were produced due to disturbance by humans or vehicles, six (10.7%) due to disturbance by non-human species, three (5.4%) by matriarchs to assemble the group ("let's go" rumble), three (5.4%) in the context of contacting other herd members (contact calls), and 24 (42.9%) during intra- and inter-group interactions at close range (Fig. 3).

Cluster analysis based on the distance matrix of pairwise measures of overall similarity between calls did not reveal discrete structural groups, suggesting that the variation in call structure is graded. Examination of the spectrograms, however, revealed differences between calls, particularly in the direction and extent of frequency modulation. Some calls showed an overall downward modulation of frequency [Fig. 4(A)], others showed little or no frequency modulation [Fig. 4(B)], and some showed an overall upward modulation in frequency [Fig. 4(C)]. Yet another type contained extensive frequency modulation within the call [Fig. 4(D)]. Preliminary comparisons did not reveal any particular correspondence between these features and either age-sex class or behavioral context, but the sample sizes for many of the groups are too small to permit meaningful conclusions at this stage.

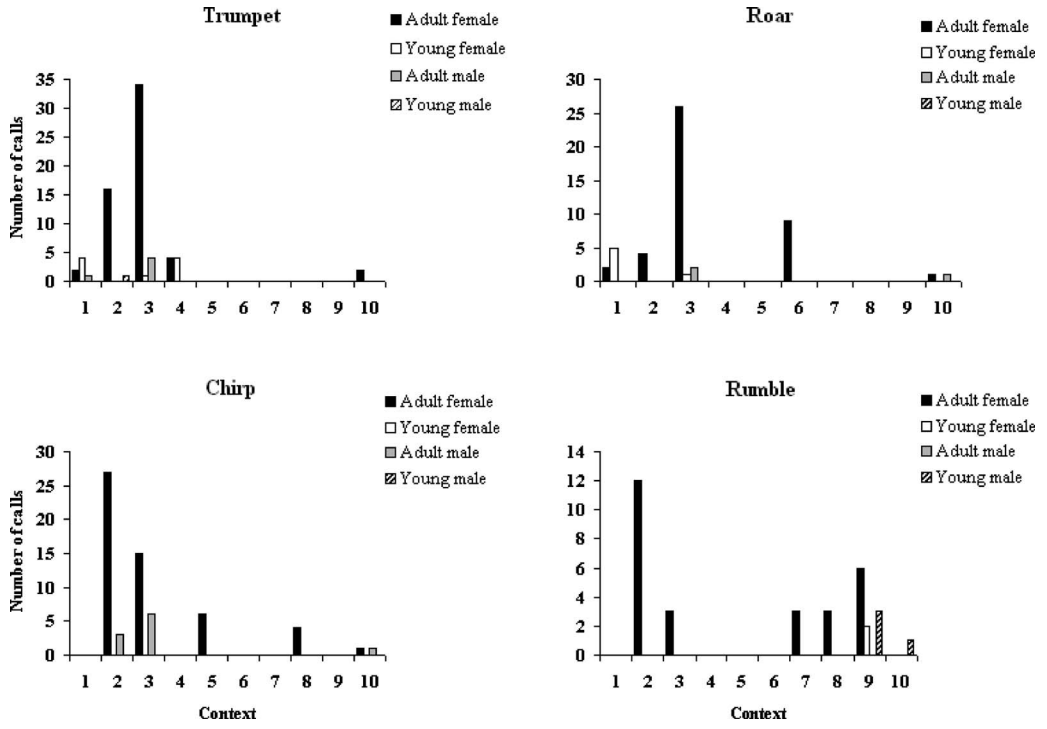


FIG. 3. Call usage according to age-sex class and social context: 1-Play, 2-disturbed by humans/vehicles, 3-disturbed by other species, 4-running out of waterhole, 5-aggression within group, 6-facing another group, 7-let's go, 8-contact call, 9-interactive calling, and 10-unknown.

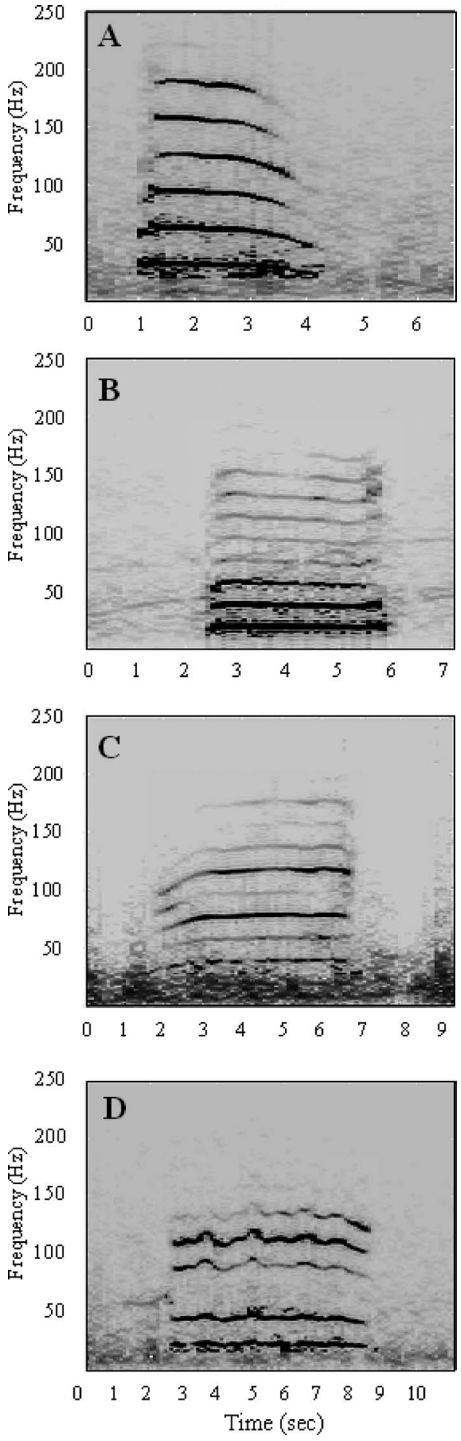


FIG. 4. Spectrograms of calls exemplifying four types of rumbles: (A) downward frequency modulation, (B) little or no frequency modulation, (C) upward frequency modulation, and (D) high modulation frequency within the call. Sampling frequency for all calls was 48 kHz with 50% overlap and window size is 200 ms.

IV. DISCUSSION

A. Structure and classification of calls

This study of the vocalizations of wild Asian elephants classifies them into four mutually exclusive categories based on structural features: trumpets, chirps, roars, and rumbles. Three of the four call types, namely, trumpets, chirps, and roars show extensive overlap in their frequency ranges. They are, however, clearly distinguishable from each other by their temporal and/or spectral structures. Trumpets show high harmonicity relative to chirps and roars, which are noisy. Chirps may be distinguished from roars by their characteristic temporal and spectral structures: short durations and frequency peaks over a narrower range. Roars exhibit low harmonicity and have no specific temporal structure. Rumbles, which constitute the fourth call type, do not overlap with any of the other call types in frequency and exhibit a distinct harmonic structure. Rumbles are also much longer in duration compared with the other call types. Our observations can be compared with previous studies on Asian and African elephants.

On the basis of auditory assessments in the field and visual assessments of a few spectrograms, McKay (1973) classified the vocalizations of wild Asian elephants in Sri Lanka and zoo elephants into three "basic sounds" with eight "resulting sounds," depending on their modification by change in amplitude, temporal patterning, and stressing of overtones, as well as non-vocal sounds produced in the trunk. However, the spectral and temporal characteristics were not defined. These "basic sounds" (with "resulting sounds") were growl (growl, rumble, roar, and "motorcycle"), squeak (chirp and trumpet), and snort ("snort" and "boom"). It is now clear that the categories of resulting sounds such as rumbles and motorcycle with infrasonic components are structurally different from roars, which are calls with low harmonicity and no infrasonic frequencies. Similarly, the chirp and the trumpet are sufficiently different in their spectrograms not to be placed together under the basic sound squeak. Non-vocal sounds such as snort and boom are not considered here since our study was confined to vocalizations. Several observers including Sanderson (1878), Krishnan (1972) and McKay (1973) also described calls of Asian elephants that clearly indicate low frequency sounds. A study on captive female Asian elephants by Payne *et al.* (1986) later recorded infrasound with fundamental frequencies of 14–24 Hz and 10–15 s duration.

There have been a number of attempts at classifying African elephant vocalizations, with most of the studies focusing on infrasound. One of the earliest studies on the African species (Berg, 1983) described the characteristics of the vocalizations of a group of captive elephants based on visual inspection of spectrograms and divided them into ten call types. Our recordings of Asian elephant vocalizations show correspondence with three of these ten types, namely, trumpets, roars, and barks [which we refer to as chirps in accordance with McKay (1973)]. The trumpets and roars recorded by Berg (1983) are longer in duration (2s and 3.8s) but similar in terms of dominant frequencies to our recordings (Table V). On the other hand, Poole *et al.* (1988) classified the low-frequency calls (rumbles) of wild African elephants into seven types based on social context. The fundamental frequencies of the Asian elephant contact calls (21.2 Hz) and let's go rumbles (18.5 Hz) are comparable to those reported by Poole *et al.* (1988) for the African elephant (Table V). More recently, Leong *et al.* (2003) provided a quantitative framework for the classification of the acoustic repertoire of captive African elephants. They described eight categories of calls, namely, trumpet, snort, croak, rev, chuff, noisy rumble,

loud rumble, and rumble. Calls similar to the noisy rumbles and loud rumbles described by Leong *et al.* (2003) were not recorded in this study. The rumbles recorded in this study correspond to their third category of rumbles where the maximum frequency is below 250 Hz. The Asian elephant trumpets, on the other hand, lie in a higher frequency range (approximately 400 Hz–6000 Hz) than those reported by Leong *et al.* (2003) for captive African elephants (Table V). Interestingly, they did not report the occurrence of two commonly observed call types, namely, roars and chirps, possibly because their study was carried out on captive elephants. Comparisons of different studies of elephant acoustic communication should take into account possible differences between free-ranging and captive animals, especially those in confined spaces as in zoos. Although only four call types are reported in this study, which was confined to vocalizations, there may exist other types of calls that were not represented in our recordings.

Elephant rumbles clearly show the highest structural diversity and attempts have been made to classify them into subtypes based on quantitative analyses of a number of acoustic features. Using a spectral cross-correlation analysis on the F_0 contour, Leong *et al.* (2003) classified rumbles into five groups, which differ in the extent of frequency modulation and duration. Soltis *et al.* (2005) failed to find distinct rumble subtypes in their study on captive African elephants. On the other hand, Wood *et al.* (2005) classified rumbles of wild African savannah elephants into three types based on the profile of the second harmonic. These differ primarily in the duration and extent of frequency modulation. Our analyses of Asian elephant rumbles, based on their spectral and temporal characteristics, did not show discrete clusters based on measures of overall structural similarity, suggesting that the variation in call structure is graded. Another possibility is that most of the variation occurs in only a few structural features and this is not captured by measures of overall similarity such as the Euclidean distance. The fact that distinct patterns of frequency modulation are clearly visible in spectrograms suggests that this may be the case.

B. Social context

Approximately 80% of the calls across all call types were made by adult or sub-adult females. Juveniles vocalized mostly in the context of play or distress. Both male and female juveniles produced three of the four call types, namely, trumpets, roars, and rumbles. They did not, however, produce chirps. The frequency of adult male vocalizations is low in our sample, but this should be interpreted with caution, since the numbers of adult males that were encountered was low.

Each of the four call types was produced in a variety of contexts and multiple call types were observed in any given context. The three major contexts in which trumpeting was observed are play (largely in the younger age classes), disturbance by humans or other species, and aggression (while charging individuals of other species or vehicles). Chirping was observed in groups that were confused or alarmed by the presence of other species (predators or otherwise) or vehicles. In the former context, the calling individuals were apparently able to detect the presence of other species through smell and they often lifted their trunks in the general direction of the source. This call typically elicited confused running and/or bunching among the other members of the herd. Another context in which this call was observed was when individuals were separated from their herds. Chirps thus seem to be associated with a state of distress or conflict within an individual.

Elephants roared when a herd first arrived at a location such as a waterhole or upon the arrival of a new herd into the area. Further, the calls were also used in the context of play and presence of other species or vehicles. Additionally, individuals, both males and females, were observed to make these calls during aggressive interactions.

Elephants produced rumbles in a variety of contexts that included interactions within and between herds and during encounters with other species. Rumbling was observed in three of the seven contexts described by Poole *et al.* (1988), namely, let's go, contact calling, and greeting. Adult female members rumbled to assemble the herd while leaving a waterhole (let's go rumble) (Poole *et al.*, 1988) or in situations of disorder and confusion such as encounters with other species or vehicles. Rumbling was observed when herd members were separated from each other and these were probably contact calls. Rumbles were also produced when two or more herds came in contact with one another and in these situations, calls were made in quick succession or simultaneously by multiple individuals, sometimes followed by trunk twining, touching, and sniffing between members of the two herds. Occasionally, adult females rumbled at juveniles involved in aggressive play. During encounters with other species (such as bears and dholes), multiple individuals rumbled simultaneously and repeatedly.

The other four call types described by Poole *et al.* (1988) are in the context of mating behavior, including the estrous rumble by females and the musth rumble by males, of which no recordings were made during the course of our study. Adult male elephants were encountered infrequently at Mudumalai due to high levels of ivory poaching in this population during the 1980s and 1990s (Arivazhagan and Sukumar, 2005). Males that were encountered, including those in musth, rarely vocalized.

At Mudumalai, rumbles were commonly produced in situations of distress and aggression. In contrast to the observations at Kruger, South Africa by Wood *et al.* (2005), no rumbles were observed in the contexts of feeding and resting unless the elephants were disturbed by other species. Trumpets, chirps, and roars were often observed to occur in overlapping contexts, including play, aggression, and disturbance. Chirps and rumbles, however, were not observed during play.

V. CONCLUSIONS

A preliminary characterization of the Asian elephant vocal repertoire has been presented, which classifies calls into four call types. Our sample sizes for the different call types are relatively small; this was on account of the often limited visibility within forests and frequent and unpredictable

movement of the herds. The sample was also biased heavily toward adult females, with relatively few calls from adult males and juveniles. Comparisons between the calls of different age-sex classes revealed no significant differences in most call features, which was surprising. This will have to be investigated further with larger samples of male and juvenile calls. Although measuring vocalizations of captive elephants can yield large sample sizes in terms of numbers of calls, the social contexts, relative frequencies of call types or even the structures themselves may deviate from those of natural populations due to the confinement, limited living space, and artificial social structure. The data in this study provide a valuable baseline since it was carried out on free-ranging elephants in the wild. Further studies are required to gain more insight into the full extent of the acoustic repertoire and the relations between call structures and social contexts in Asian elephants.

ACKNOWLEDGMENTS

This research was funded by the Department of Science and Technology, Government of India. Permits were provided by the Tamilnadu Forest Department to carry out research in Mudumalai Wildlife Sanctuary. Mr. Maran and Mr. Kunmari are acknowledged for their assistance in fieldwork.

Arivazhagan, C., and Sukumar, R. (**2005**). "Comparative demography of Asian elephant populations (*Elephas maximus*) in southern India," CES Technical Report No. 106, Centre for Ecological Sciences, Indian Institute of Science, Bangalore.

Berg, J. K. (**1983**). "Vocalizations and associated behaviours of the African elephant (*Loxodonta africana*) in captivity," Z. Tierpsychol **63**, 63–79.

Boll, S. F. (**1979**). "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process. **27**, 113–120.

Bradbury, J. W., and Vehrencamp, S. L. (**1998**). *Principles of Animal Communication* (Sinauer Associates, Inc., Sunderland, MA).

Clemins, P. J., Johnson, M. T., Leong, K. M., and Savage, A. (**2005**). "Automatic classification and speaker identification of African elephant (*Loxodonta africana*) vocalizations," J. Acoust. Soc. Am. **117**, 956–963.

Douglas-Hamilton, I. (**1972**). "On the ecology and behaviour of the African elephant," Ph.D. thesis, University of Oxford, Oxford, UK.

Garstang, M., Larom, D., Raspet, R., and Lindeque, M. (**1995**). "Atmospheric controls on elephant communication," J. Exp. Biol. **198**, 939–951.

Krishnan, M. (**1972**). "An ecological survey of the larger mammals of peninsular India: The Indian elephant," J. Bombay Nat. Hist. Soc. **69**, 469–501.

Langbauer, W. R., Jr. (**2000**). "Elephant communication," Zoo Biol. **19**, 425–445.

Langbauer, W. R., Jr., Payne, K. B., Charif, R. A., Rapaport, L., and Osborn, F. (**1991**). "African elephants respond to distant playbacks of low-frequency conspecific calls," J. Exp. Biol. **157**, 35–46.

Larom, D., Garstang, M., Lindeque, M., Raspet, R., Zunckel, M., Hong, Y., Brassel, K., O'Beirne, S., and Sokolic, F. (**1997**). "Meteorology and elephant infrasound at Etosha National Park, Namibia," J. Acoust. Soc. Am. **101**, 1710–1717.

Leighty, K. A., Soltis, J., Wesolek, C. M., and Savage, A. (**2008**). "Rumble vocalizations mediate interpartner distance in African elephants, *Loxodonta africana*," Anim. Behav. **76**, 1601–1608.

Leong, K. M., Ortolani, A., Burks, K. D., Mellen, J. D., and Savage, A. (**2003**). "Quantifying acoustic and temporal characteristics for a group of captive African elephants *Loxodonta africana*," Bioacoustics **13**, 213–231.

Maglio, V. J. (**1973**). "Origin and evolution of the Elephantidae," Trans. Am. Phil. Soc. **63**, 1–149.

Manly, B. F. J. (**1986**). *Multivariate Statistical Methods* (Chapman and Hall, London).

McComb, K., Moss, C. J., Durant, S. M., Baker, L., and Sayialel, S. (**2001**). "Matriarchs are repositories of social knowledge in African elephants," Science **292**, 491–494.

McComb, K., Moss, C. J., Sayialel, S., and Baker, L. (**2000**). "Unusually extensive networks of vocal recognition in African elephants," Anim. Behav. **59**, 1103–1109.

McComb, K., Reby, D., Baker, L., Moss, C. J., and Sayialel, S. (**2003**). "Long distance communication of acoustic cues to social identity in African elephants," Anim. Behav. **65**, 317–329.

McKay, G. M. (**1973**). "Behaviour and ecology of Asiatic elephants in southeastern Ceylon," Smithson. Contrib. Zool. **125**, 67–69.

Moss, C. J. (**1988**). *Elephant Memories* (Elm Tree, London).

Moss, C. J., and Poole, J. H. (**1983**). "Relationship and social structure in African elephants," in *Primate Social Relationships: An Integrated Approach*, edited by R. A. Hinde (Blackwell Scientific, Oxford), pp. 315–325.

Oppenheim, A. V., and Schafer, R. W. (**1989**). *Discrete-Time Signal Processing* (PrenticeHall, Princeton, NJ).

Payne, K. (**1998**). *Silent Thunder: The Hidden Voice of Elephants*, 2nd ed. (Phoenix, London).

Payne, K. B., Langbauer, W. R., Jr., and Thomas, E. M. (**1986**). "Infrasonic calls of the Asian elephant (*Elephas maximus*)," Behav. Ecol. Sociobiol. **18**, 297–301.

Poole, J. H. (**1994**). "Sex differences in the behaviour of African elephants," in *The Differences Between the Sexes*, edited by R. V. Short and E. Balaban (Cambridge University Press, Cambridge), pp. 331–346.

Poole, J. H. (**1999**). "Signals and assessment in African elephants: Evidence from playback experiments," Anim. Behav. **58**, 185–193.

Poole, J. H., and Moss, C. J. (**1989**). "Elephant mate searching: Group dynamics and vocal and olfactory communication," in *The Biology of Large African Mammals in their Environment*, edited by P. A. Jewell and G. M. O. Maloiv (Clarendon Press, Oxford), pp. 111–125.

Poole, J. H., Payne, K., Langbauer, W. R., Jr., and Moss, C. J. (**1988**). "The social context of some very low frequency calls of African elephants," Behav. Ecol. Sociobiol. **22**, 385–392.

Sanderson, G. P. (**1878**). *Thirteen Years Among the Wild Beasts of India* (W.H. Allen, London).

Sneath, P. H. A., and Sokal, R. R. (**1973**). *Numerical Taxonomy* (W.H. Freeman and Company, San Francisco, CA).

Soltis, J., Leong, K., and Savage, A. (**2005**). "African elephant vocal communication II: Rumble variation reflects the individual identity and emotional state of callers," Anim. Behav. **70**, 589–599.

Stoeger-Horwath, A. S., Stoeger, S., and Schwammer, H. M. (**2007**). "Call repertoire of infant African elephants: First insight into the early vocal ontogeny," J. Acoust. Soc. Am. **121**, 3922–3931.

Sukumar, R. (**1989**). *The Asian Elephant: Ecology and Management* (Cambridge University Press, Cambridge).

Sukumar, R. (**2003**). *The Living Elephants: Evolutionary Ecology, Behavior and Conservation* (Oxford University Press, New York).

Varman, K. S., and Sukumar, R. (**1995**). "The line transect method for estimating densities of large mammals in a tropical deciduous forest: An evaluation of methods and field experiments," J. Biosci. **20**, 273–287.

Venkataraman, A. B., Venkatesa Kumar, N., Varma, S., and Sukumar, R. (**2002**). "Conservation of a flagship species: Prioritizing Asian elephant (*Elephas maximus*) conservation units in southern India," Curr. Sci. **82**, 1022–1032.

Vidya, T. N. C., and Sukumar, R. (**2005**). "Amplification success and feasibility of using microsatellite loci amplified from dung to population genetic studies of the Asian elephant (*Elephas maximus*)," Curr. Sci. **88**, 489–492.

Wittemyer, G., Douglas-Hamilton, I., and Getz, W. M. (**2005**). "The socioecology of elephants: Analysis of the processes creating multitiered social structures," Anim. Behav. **69**, 1357–1371.

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.

Effects of syllable-final segment duration on the identification of synthetic speech continua by birds and humans

Thomas E. Welch, James R. Sawusch, and Micheal L. Dent[a)]
Department of Psychology and Center for Cognitive Science, University at Buffalo, the State University of New York, Buffalo, New York 14260

(Received 18 December 2008; revised 30 July 2009; accepted 31 July 2009)

In an attempt to test whether experience with or knowledge of language is necessary to show typical speaking rate effects in the perception of speech, budgerigars (*Melopsittacus undulatus*) and humans categorized stimuli from the synthetic continua /ba/-/wa/ and /bas/-/was/, with both short and long syllable-final phonemes. This comparative approach aims to shed some light on whether knowledge of language has a role in rate normalization effects, such as using duration information as an indicator of speaking rate in human speech perception. Syllable-final phoneme durations were varied, and were either temporally adjacent to the initial target (CV series) or were nonadjacent (CVC series). The birds were always influenced by syllable-final duration variation in the present experiments and displayed greater boundary shifts than humans. In humans, there was a significant boundary shift observed in the CV series, but there were no effects of duration variation in the final segment in the CVC series. The results from the birds suggest that specialized speech-based principles may not be necessary for explaining findings of grouping speech or speechlike elements in perception. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3212923]

PACS number(s): 43.80.Lb, 43.71.Es, 43.71.Rt [MO] Pages: 2779–2787

I. INTRODUCTION

Research in speech perception is often focused on gaining an understanding of how human listeners recognize speech sounds despite tremendous variability in any given signal, and ultimately on how properties of the acoustic signal are mapped onto linguistic elements. Changes in speaking rate, talkers, dialect, and environmental noise are a few ways in which the acoustic signal can vary. Since human perception of speech sounds is known to be heavily influenced by the acoustic properties of the contexts in which the sounds are heard, perception must somehow take this variability into account, or actively exploit it, to achieve constancy.

One of the sources of variability in an acoustic signal is the speaking rate of the talker. Consequently, the physical duration of different segments of the speech signal (e.g., phonemes and syllables) changes as speaking rate varies (Crystal and House, 1982, 1990). Evidence from large speech databases has shown that speaking rate varies both across talkers (Byrd, 1992) and also dynamically within the speech of an individual talker (Miller *et al.*, 1984). This issue of rate change is particularly important because segment duration is a major acoustic cue to many phonetic distinctions. Thus, as speaking rate changes, so do the durations of acoustic segments that cue these distinctions. For instance, the /b/-/w/ stop versus semivowel manner contrast can be cued by duration differences alone, with shorter initial formant transitions heard as more "b-like" and longer transitions as more "w-like" (Miller and Liberman, 1979). When a talker is speaking very fast, listeners still hear /w/ phonemes even though the physical duration information may be the same as what would normally cue a /b/ phoneme at a slower rate of speech (Miller and Baer, 1983). Conversely, when listening to someone who speaks very slowly, the intended /b/s do not all sound like /w/s, they are often still heard as /b/.

Research examining exactly how listeners compensate for differences in speaking rate between different talkers and within any given talker seems to show that some sort of "normalization" or scaling processes are at work. For instance, using synthetic speech, Ainsworth (1973) found that when the duration of sounds in a precursor sequence was longer (specifying a slower speech tempo), a longer initial transition was required to hear a semivowel instead of a stop consonant. For faster speech, a relatively shorter transition was perceived as a semivowel and not a stop. Similar results have also been obtained using edited natural speech (e.g., Minifie *et al.*, 1976). In addition to this short term influence of prior information, further removed precursor information, such as the rate of earlier stressed syllables, can also have an effect on the perception of a later target (Kidd, 1989; Miller, 1981; Summerfield, 1981). Miller and Liberman (1979) also showed that the duration of a segment that follows the target (/a/ in a /ba/-/wa/ series) can influence listeners' perceptions of the earlier duration-based target distinction.

Other studies (Miller and Liberman, 1979; Newman and Sawusch, 1996; Sawusch and Newman, 2000; Summerfield, 1981) have examined the influence of later, nonadjacent, phonetic segment duration on perception. These studies have shown that segment duration is the critical factor in influencing listener perception, not syllable duration (Miller and Liberman, 1979), and that all segment durations within a short temporal window following the phoneme target influence perception (see Sawusch and Newman, 2000 for a summary).

[a)]Author to whom correspondence should be addressed. Electronic mail: mdent@buffalo.edu

While context-dependent perception has obvious linguistic benefits in understanding speech, it is not unique to the perception of speech. For instance, rate normalization is also found in the perception of other (nonspeech) auditory events (e.g., Diehl and Walsh, 1989; Fowler, 1990; Pisoni *et al.*, 1983). Fowler (1990) proposed that listeners' perception is sensitive to "event" rate. Support for this view is found in data from infant and nonhuman participants with very limited experience or knowledge of language. For example, prelinguistic infants demonstrate rate-dependent perception of the /b/-/w/ contrast because their ability to discriminate a difference in transition duration is dependent on the duration of a following vowel (Eimas and Miller, 1980). Similar results have been found in studies of the perception of nonspeech stimuli by infants (Jusczyk *et al.*, 1983). Moreover, budgerigars, who have the ability to mimic human speech sounds, also show rate-dependent effects (Dent *et al.*, 1997), as do Japanese macaques (*Macaca fuscata*), who cannot produce speech sounds (Sinnott *et al.*, 1998). While it is certainly true that infants are exposed to language both *in utero* and following birth, and domesticated birds are constantly hearing the chatter of human speech (but with no real functional significance) around the laboratory, it can be assumed that their experiences are rather modest when compared to the tens of thousands of hours of experience with language that adult humans possess. This issue makes the infant and animal studies, particularly those that examine possible experience-based influences on the perception of human speech, invaluable to understanding the mechanisms of auditory processing.

Another reason for testing the perception of duration-varying speech stimuli by animals, even though there are data available from humans tested on nonspeech stimuli, is to rule out the possibility that humans may be simply transferring their speech perception strategies or "modes" to nonspeech stimuli. Additionally, speech sound processing can be further examined in nonhuman animals by searching for neural correlates to the behavioral results. Moreover, birds that can produce speech offer the only known nonhuman opportunity to test the relationship between production and perception. While such comparative data do not do much to differentiate whether rate-dependent speech perception arises from general perceptual mechanisms like durational contrast (see Lotto and Kluender, 1998; Wade and Holt, 2005) or from recovery of environmental sources (i.e., articulatory gestures, see Fowler, 1990), they do indicate that significant knowledge of language, including rate-dependent speech patterns, may not be necessary for rate-dependent speech categorization. Whereas many past rate-dependent findings seem to support a more general contrastive mechanism (e.g., Dent *et al.*, 1997; Lotto and Kluender, 1998; Sinnott *et al.*, 1998), there certainly exists a need to more fully explore both the task and stimulus properties that seem to be influencing such effects.

In the present study, the locations and shifts of stop-semivowel phoneme boundary placements were investigated in budgerigars and humans, two species with similar psychoacoustic capabilities (see Dooling *et al.*, 2000). Budgerigars previously demonstrated sensitivity to the duration of the vowel following a /b/-/w/ contrast using synthetic speech stimuli and discrimination procedures (Dent *et al.*, 1997). Unfortunately, while discrimination procedures were used to test the effect in birds, identification procedures were used to test the human subjects in the study of Dent *et al.* (1997) on budgerigars. Sinnott *et al.* (1998) also used identification procedures in a macaque study using similar stimuli. The present study aims to obtain identification data from birds that could be compared more directly with the human and macaque data.

In Experiment 1, two /ba/-/wa/ series similar to those used by Dent *et al.* (1997) and Miller and Liberman (1979) were used. By varying the duration of the syllable-final vowel in short and long synthetic /ba/-/wa/ stimuli and measuring categorization of these stimuli, it is possible to see if avian category and human phoneme boundaries shift toward longer transition durations as a function of increasing vowel length. In Experiment 2, we examined the duration-based stop-semivowel context effect using a duration-varying, temporally nonadjacent final /s/ fricative. The influences of an additional, temporally removed, noise-based fricative were investigated because of what is known about the temporal limitations of normalization effects in humans (e.g., Newman and Sawusch, 1996) and the principles of perceptual grouping (e.g., Bregman, 1990) as they apply to speech (Remez *et al.*, 1994).

In the /bas/-/was/ study, the stimuli were synthetically generated and stylized as in the original Miller and Liberman (1979) experiments. Newman and Sawusch (1996) postulated that the time course of acoustic-phonetic processing (e.g., size of the temporal window) varies as a function of stimulus quality. With high-quality stimuli like the edited natural speech used by Newman and Sawusch (1996), it was proposed that phonetic processing happens relatively quickly, and thus there might be a shorter temporal window available for subsequent segment duration to influence perception (~250 ms from target onset). With lower-quality stimuli such as the highly stylized synthetic tokens used by Miller and Liberman (1979), perceptual processing may take longer before a phonetic decision can be made. This relatively longer temporal window for lower-quality stimuli (within which segments of a speech stream can influence rate normalization) may be a possible explanation for Miller and Liberman's (1979) normalization effects for information ranging up to 400 ms from target onset. Since the stimuli in the current study were synthetically generated and stylized, we predicted normalization effects (at least for humans) as long as the duration-varying information fell within about 350–400 ms from the onset of the target. While a lack of an effect does not necessarily rule out a temporal window hypothesis, it would perhaps raise some interesting questions about stimulus characteristics and quality, which may be affecting this dynamic window of influence. To the extent that the changes in human classification of the initial /b/-/w/ contrast are rooted in auditory processes of durational contrast, we would expect that the birds would also be influenced by the duration of a nonadjacent segment.

The present experiments were also designed to identify other general and speech-specific principles that may be af-

fecting perceptual grouping of the acoustic information in the /bas/-/was/ syllables. The experiments test whether the acoustic elements in the speech signals will bind as one coherent stream or divide into separate streams or auditory events. If the duration of the later-occurring fricative /s/ in the /bas/-/was/ sets influences the perception of the syllable-initial stop-semivowel contrast, it would suggest that the fricative was perceived as belonging to a unitary stream with the initial consonant and vowel. An /s/ was chosen for the final consonant in part because fricatives vary naturally in duration as a function of speaking rate (Crystal and House, 1982, 1988). In addition, the /s/ noise-based fricative is aperiodic unlike the relatively periodic initial /b/-/w/ contrast and intervening vowel segment. This spectral dissimilarity of sounds within the signal functionally tests the Gestalt *similarity* principle of perceptual grouping in that it would be expected that the /s/ should be less likely to cohere to the initial segments and influence the perception of the earlier /b/-/w/ target (Bregman, 1990).

With humans, it has been shown that dissimilar phonemes and phonemes produced by different talkers can influence the perception of the initial target phoneme so long as the rate-bearing information falls within the limited temporal window (e.g., Sawusch and Newman, 2000). This result is in accord with the Gestalt organizing principle of *proximity*. In spite of differences in spectral quality and a lack of good-continuation between the /ba/-/wa/ CV and the final /s/, close temporal proximity or contiguity of the final /s/ to the to-be judged CV should still ensure the binding of the CV and /s/ into one stream (e.g., Bregman, 1990). Thus, with respect to Gestalt principles of perceptual organization, the final /s/ may either cohere with the initial CV (contiguity) or separate from it (differences in periodicity, minimal spectral continuity).

These experiments also aim to investigate whether knowledge of language is needed for the speech stimuli to cohere or bind into a single sound. For instance, Remez *et al.* (1994), based on experiments using sinewave replicas of utterances, proposed speech-specific principles of coherence that can be used to group disparate signal elements together, even when the basic auditory principles listed by Bregman (1990) appear to fail. This issue is relevant to Experiment 2 in the current study, where a nonadjacent and aperiodic (noisy) phoneme is added to the earlier-occurring periodic acoustic elements. Phonetic coherence, as laid out by Remez *et al.* (1994), suggests that when basic Gestalt grouping principles like similarity fail to explain the observed cohesion of elements in an auditory stream, there may be additional sensitivities to acoustic modulation characteristic of speech signals that humans use in perceptual organization. Humans may have a special advantage in grouping otherwise disparate acoustic elements and may be relying on a principle of grouping that is sensitive to the acoustic products of vocalizations (see Remez *et al.*, 1994). If this specialization is the case, human listeners with knowledge of the language may be able to utilize this "phonetic coherence" in order to group the final /s/ to the /ba/-/wa/ base, despite the clear dissimilarity in periodicity.

For humans, it would seem that the primitive Gestalt-based auditory grouping principle of temporal proximity along with the idea of phonetic coherence proposed by Remez *et al.* (1994) should both be acting to bind all of the information in these brief synthetic utterances. Since birds possess no (or minimal) knowledge of the human vocal tract, it would suggest that if they show an effect of the final aperiodic fricative on their boundary placement, that phonetic coherence is not necessary to explain the grouping. This outcome would in turn suggest that phonetic coherence may not be necessary for humans either, and that close temporal contiguity alone may be sufficient for the formation of a single speech stream. If birds fail to show a syllable-final effect for Experiment 2, it follows that something other than basic Gestalt principles are needed to bind the diverse acoustic elements into one stream, consistent with the work of Remez *et al.* (1994). Experiment 2 investigates whether phonetic coherence is necessary to bind the diverse acoustic elements of speech signals together by comparing the performance of humans and birds with the /bas/-/was/ series where the /s/ was short or long in duration.

II. METHODS

A. Subjects

Five adult budgerigars (four males and one female) were used as subjects in these experiments. Each bird ran in both experiments. All of the birds were individually housed in a vivarium at the University at Buffalo, SUNY and were kept on a day/night cycle corresponding to the season. The birds were either purchased from a local pet store or bred in the vivarium. They were food restricted to approximately 90% of their free-feeding weight during the course of the experiment. All procedures were approved by the University at Buffalo, SUNY's Institutional Animal Care and Use Committee and complied with NIH guidelines for animal use.

Forty-eight undergraduate listeners from the University at Buffalo, SUNY also participated in the experiments for course credit. Eighteen participants listened to the two sets of stimuli that varied in final /a/ duration (/ba/-/wa/ series), and 30 heard the sets that were varied in the final /s/ duration (/bas/-/was/ series).

B. Testing apparatus

The psychoacoustic experiments in birds took place in one of four identical psychoacoustic testing setups. The setups consisted of a wire test cage ($61 \times 33 \times 36$ cm^3) mounted in a sound-attenuated chamber (Industrial Acoustics Co., Small Animal Chamber) lined with sound-absorbent foam (10.2 cm Sonex, Ilbruck Co.). The test cage consisted of a perch, an automatic food hopper (Med Associates Standard Pigeon Grain Hopper), and two vertical response keys extending downwards from the inside of the hopper in front of the bird. The response keys were two sensitive microswitches with 1 cm^2 green (left key) or red (right key) buttons glued to the ends. The birds pecked the colored keys, which tripped the microswitches. A 7-W light at the top of the test cage illuminated the chamber and served as the experimental house light. An additional 30-W bulb remained

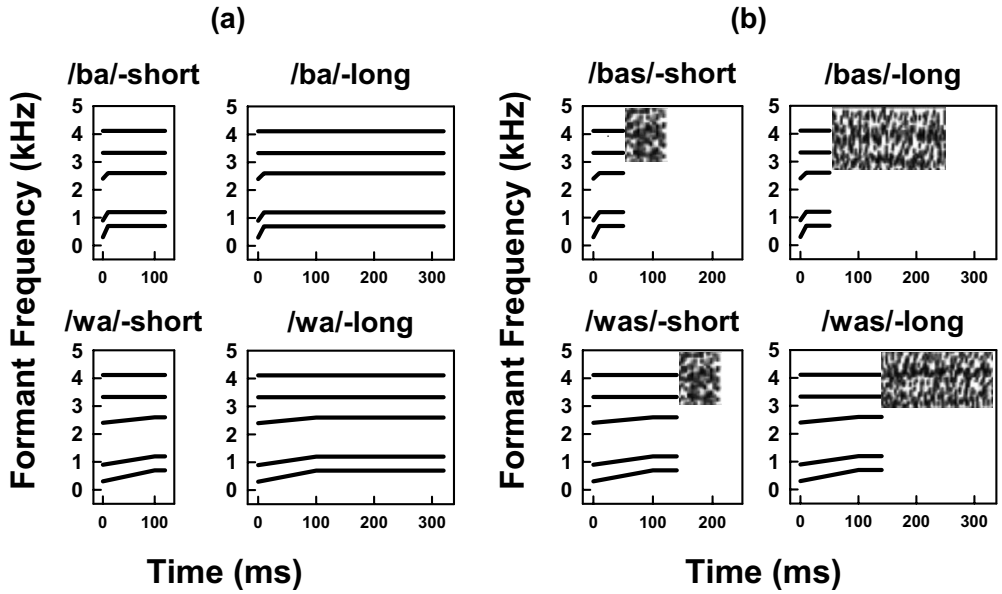


FIG. 1. Schematics of the short and long end point stimuli for both the /ba/-/wa/ and /bas/-/was/ sets. (a) Short /ba/-/wa/ continuum of 122 ms and long /ba/-/wa/ continuum of 305 ms. (b) Short /bas/-/was/ continua always has 80 ms of final /s/ frication (130 ms and 220 ms total duration for each respective end point), and long /bas/-/was/ continua always has 200 ms of final /s/ frication (250 and 340 ms total duration for each respective end point).

on in the chamber for the entire session. The behavior of the animals during test sessions was monitored at all times by an overhead web camera (Logitech QuickCam Pro, model 4000). One speaker (Morel Acoustics, model MDT-29) was hung directly behind the subject at the level of the bird's head, 30.5 cm away from the bird during testing. The experiments were controlled by a Dell microcomputer (Dimension 4600) operating Tucker-Davis Technologies modules (TDT, Gainesville, FL; Models RX6, PA5, SA1 and MS2) and SYKOFIZX psychophysics software.

Human listeners were run individually and stimulus presentation and responses were controlled by an Apple Macintosh (G4) computer. The stimuli were presented at a 10 kHz sampling rate with 16 bits/sample, amplified and played binaurally through TDH-39 headphones. Listeners were asked to indicate whether each syllable began with a /b/ or a /w/ by pressing the appropriate key on a computer-controlled keyboard.

C. Stimuli

All stimuli were generated in their entirety using the cascade mode of Klatt's (1980) cascade/parallel synthesizer software. The /ba/-/wa/ and /bas/-/was/ continua used in the experiments were five-formant speech tokens that each had ten items that began with a phoneme ranging from /b/ to /w/. Two sets were /ba/-/wa/ series that differed only in duration of the ending vowel (long and short series) and the remaining two continua were /bas/-/was/ sets that differed only in duration of the syllable-final /s/ fricative. The initial /b/-/w/ phonemes for each continuum were created by increasing the duration of the formant transitions in 10-ms steps from 10 to 100 ms to yield series of ten stimuli. Additionally, the duration of the amplitude rise time at syllable onset was varied systematically to produce more natural sounding synthetic tokens.

Figure 1 shows a schematic representation of the end point stimuli with both long and short syllable-final phonemes. For all continua, F1 began at 300 Hz and moved to 700 Hz over a variable time period from 10 ms (/ba/ end) to 100 ms (/wa/ end). F2 moved from 900 to 1200 Hz, and F3 moved from 2300 to 2600 Hz over the same variable time period. F0 fell linearly over the duration of the syllable from 125 to 80 Hz. F4 and F5 remained fixed at 3300 and 4100 Hz. The bandwidths for F1–F5 were 90, 90, 130, 400, and 500 Hz respectively. The vowel /a/ was steady state for all stimuli. For the /ba/-/wa/ stimuli, the duration of the steady-state vowel was shortened as the transition duration was lengthened to create one set of short syllables (122 ms overall duration) and one set of long syllables (305 ms overall duration).

For the /bas/-/was/ stimuli, the duration of the vowel was fixed (40 ms) while the duration of the final /s/ fricative remained at 80 ms for the short series and 200 ms for the long series. The frication was created by slowly turning the amplitude of frication on and then turning off the amplitude of voicing after 10 ms of frication. The amplitude profile for the /s/ involved setting A4 to 24 dB, A5 to 40 dB, and A1–A3 to 0 dB. ADOBE AUDITION (version 1.5) and PRAAT (version 5.1.07) were used to inspect and verify the construction of the stimuli.

D. Procedures

Birds. The birds were randomly assigned to either the /ba/-/wa/ or /bas/-/was/ short or long series and were trained to differentially respond to the end point stimuli at a set criterion level of performance before moving onto testing. The birds were first trained using operant conditioning procedures to peck the keys for food reinforcement. To start a trial, the birds first pecked the left key, which would activate a variable interval of 2–7 s. At the end of the variable inter-

val, one of two end point speech sounds (i.e., beginning with a 10 ms transition /b/ or a 100 ms transition /w/) was presented with equal probability. The birds were trained to peck the left key again when they heard a sound beginning with /b/ and to peck the right key when they heard stimuli beginning with /w/. The entire stimulus sound played, regardless of when the animal responded. If they correctly identified the sound within 2 s of the beginning of the stimulus, they were rewarded with 1.5 s access to hulled millet from the illuminated food hopper for 70% of the correct trials. They were rewarded with the hopper light only for 1.5 s in the other 30% of correct trials, signifying a correct response but keeping the birds motivated to run more trials. If the response was longer than 2 s or if the animal did not respond at all, no reward was given. If they responded incorrectly by indicating the wrong category of the stimulus (e.g., a left peck for a /w/ sound), the house light was extinguished for 5 s. As soon as the reinforcement or punishment phases were completed, the animals could immediately initiate another trial by pecking the left key again.

During training, birds were only presented with the two end point /b/ and /w/ (10 and 100 ms transitions) stimuli from the series. Percent-correct scores were calculated for 100-trial blocks and the birds were moved onto testing, where the intermediate stimuli in the series were introduced, when they scored 85% correct or better on three successive 100-trial blocks. Similarly, to increase comparability, only human data where listeners were above 80% on identifying end points were considered.

During the testing phase, intermediate stimuli in the series (stimuli 2–9) were randomly interspersed with the end point stimuli on 30% of trials. The responses to intermediate stimuli were always rewarded to minimize any biased responding and to maintain stimulus control. The percentage of /b/ responses was calculated for all birds for each token in the four continua. All four series were tested in a random order and a different random order was used for each bird. Each bird ran at least 533 trials on each of the four continua, resulting in 20 presentations of each stimulus type.

Humans. Listeners were run individually. Each listener was presented with either short and long /ba/-/wa/ or short and long /bas/-/was/ continua. Syllables were presented in a random order over headphones. Listeners made a two alternative forced choice (/b/ or /w/) for each syllable. The presentation rate depended on the listeners' response speed and the next syllable was presented as soon as the listener responded. No feedback was provided. Responses from the first block of 60 trials were considered practice and were not included in the analysis. After the practice block, stimuli were presented in four blocks of 80 trials, representing four repetitions of each of the 20 stimuli per block. The final analyses were based on 16 presentations of each stimulus per listener.

E. Data analysis

Analysis of the experimental data for humans and birds consisted of computing a mean percentage of /b/ responses given by all listeners to each stimulus in every series presented. The /b/-/w/ category boundary for each listener was then determined for each series by linear interpolation of the transition duration corresponding to the 50% /b/ response point. Paired sample t-tests were used to determine if identification functions changed as a function of syllable-final segment duration for each species. One-way analyses of variance (ANOVAs) were used to determine if there were boundary shift differences between experiments and species. Boundary locations were also analyzed with a two-way ANOVA to test for species and series differences for each experiment.

TABLE I. Category boundary locations (transition duration in ms) for the /a/ vowel varying /ba/-/wa/ series in Experiment 1 for birds (top left) and humans (top right). The bottom rows show boundary locations for the /s/ fricative varying /bas/-/was/ series in Experiment 2 for birds (bottom left) and humans (bottom right).

Boundary location (ms)					
		Birds		Humans	
		M	SD	*M*	SD
/ba/-/wa/	Short series	39	8.64	39	3.96
	Long series	60	6.88	45	6.33
/bas/-/was/	Short series	43	8.65	45	5.25
	Long series	57	11.60	46	6.69

III. RESULTS

A. Experiment 1: Syllable-final /a/ duration varied

Table I shows the mean boundary locations and standard deviations for the /ba/-/wa/ (/a/ varying) sets collapsed across five birds and 18 human listeners. The boundaries for human listeners were at 39 ms for the short series and 45 ms for the long series, yielding a mean boundary shift of 6 ms. For birds, mean boundaries were located at 39 and 60 ms for short and long continua, respectively, producing a mean shift of 21 ms. Figure 2 shows the mean identification functions for both bird and human listeners on these adjacent vowel duration-varying sets.

The difference between the mean boundaries for the short- and long-vowel duration series was significant for both humans $[t(17)=5.54,\ p<0.001]$ and birds $[t(4)=9.61,\ p<0.001]$. Although both species exhibited a significant overall boundary shift toward longer transition durations as a function of increasing the duration of the vowel, birds displayed a significantly larger boundary shift than humans $[F(1,21)=29.06,\ p<0.001]$, as calculated by a one-way ANOVA.

Boundary locations were also analyzed with a two-factor ANOVA and it was found that both main effects were significant: species $[F(1,42)=17.38,\ p<0.001]$, indicating that bird boundaries were at longer transition durations than human boundaries and series $[F(1,42)=34.94,\ p<0.001]$, indicating that the boundaries shifted from the short to the long series. The interaction was also significant $[F(1,42)=9.71,\ p<0.005]$, and Holm–Sidak *post hoc* tests revealed that this result was due to the significant species difference in

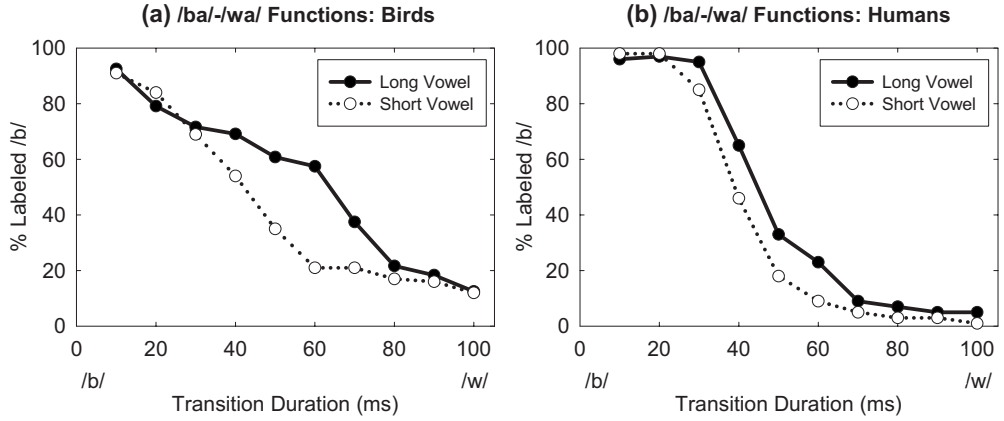


FIG. 2. Mean identification data for five bird (a) and 18 human (b) listeners for the adjacent vowel varying condition (/ba/-/wa/ sets). Percentages of /b/ responses are plotted as a function of initial transition duration in ms for the short and long syllable (vowel) series.

the long-vowel condition ($p<0.05$). The results from this experiment show that (1) the category boundary for both humans and birds shifted with changes in the duration of the adjacent vowel, (2) birds exhibited greater boundary shifts than humans, and (3) that the bird boundaries were at longer durations than those for the humans in the long-vowel condition.

B. Experiment 2: Syllable-final /s/ duration varied

Table I shows the mean boundary locations and standard deviations for the /bas/-/was/ (/s/ varying) sets collapsed across five birds and 30 human listeners. Figure 3 shows the mean identification functions for both bird and human listeners on the nonadjacent fricative duration-varying sets. For birds, mean boundaries were located at 43 and 57 ms for long and short continua, respectively, producing a mean shift of 14 ms. The boundaries for human listeners were 45 ms for the short series and 46 ms for the long series.

The difference between the mean boundaries for the short and long frication duration series was significant for birds $[t(4)=7.04,\ p<0.005]$ but not for humans $[t(29)=1.23,\ p>0.05]$. Boundary locations were also analyzed with a two-factor ANOVA to examine the effects of *species* and *series*. The main effect of species was not significant $[F(1,66)=2.10,\ p>0.05]$, indicating that mean bird and human boundary locations did not differ, but the main effect of series was significant $[F(1,66)=10.35,\ p<0.005]$, indicating that the overall boundaries shifted from the short to the long series. The interaction was also significant $[F(1,66)=7.84,\ p<0.05]$, due to the contribution of the significant boundary shift from short to long duration series for budgerigars ($p<0.05$). Boundary shift differences were also compared between Experiments 1 and 2 for each species using one-way ANOVAs and they were found to be nonsignificant for birds $[F(1,8)=4.64,\ p>0.05]$ but significant for humans $[F(1,46)=18.03,\ p<0.001]$ (Fig. 4). Birds exhibited similar boundary placements and shifts across the two experiments. Human boundaries did not move significantly in Experiment 2 as a function of variation in the nonadjacent fricative /s/ like they did in Experiment 1 for the adjacent vowel duration variation.

Figure 5 shows category boundary and shift data for /b/-/w/ continua from the current identification experiments, budgerigar discrimination data, and human identification data from the budgerigar study of Dent *et al.* (1997), macaque and human identification data from the study of Sinnott *et al.* (1998), and human identification data from the classic study of Miller and Liberman (1979). While the procedures, stimuli, and species all varied across studies, boundary locations and shifts with syllable duration are quite simi-

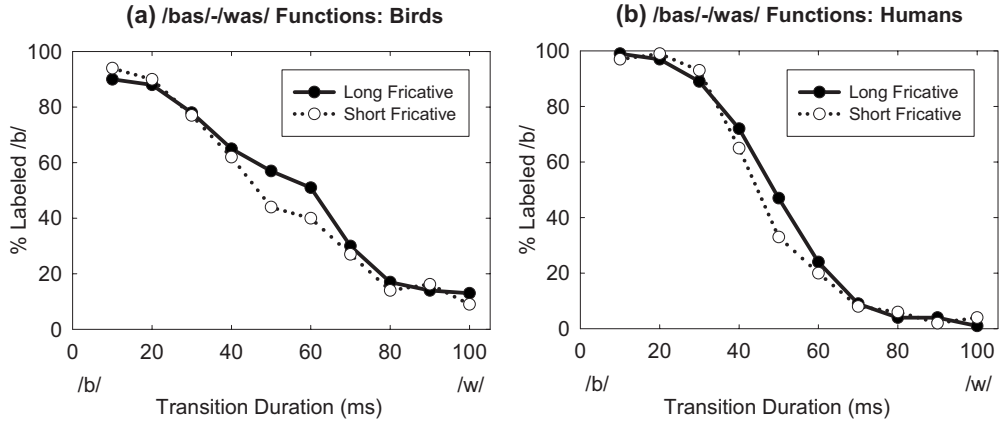


FIG. 3. Mean identification data for five bird (a) and 30 human (b) listeners for the nonadjacent fricative varying condition (/bas/-/was/ sets). Percentages of /b/ responses are plotted as a function of initial transition duration in ms for the short (80 ms) and long (200 ms) /s/ series.

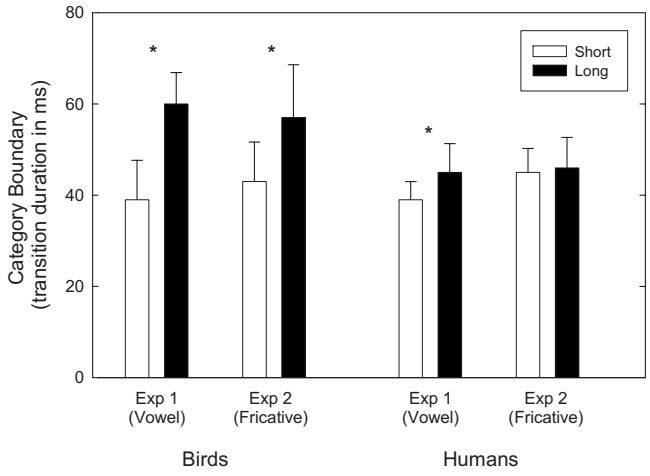


FIG. 4. Mean /b/-/w/ category boundaries (transition duration in ms) for humans and birds for Experiment 1 (/ba/-/wa/ series) and Experiment 2 (/bas/-/was/ series). Error bars represent within-condition standard deviations. All significant category shifts ($p < 0.05$) are indicated by asterisks.

lar.

IV. DISCUSSION

The present study demonstrates that later-occurring speech events can influence the identification of preceding speech events, even across species and contexts. The purpose of the present experiments was to use identification procedures to compare birds and humans in a well-known context effect that occurs with speech sounds: the influence of speaking rate on the stop-semivowel phoneme boundary. The extent to which the listeners were adjusting for rate was inferred from their differences in labeling stimuli as /b/ or /w/ as a function of the duration of a later-occurring segment. An additional goal was to use stimuli that were designed to fit within the known temporal limitations of these normalization effects to then test principles of grouping relating to temporal proximity and spectrotemporal similarity. Since birds have no knowledge of language and humans certainly do, the species comparisons allow us to make inferences about the role of specialized knowledge or processes specific to speech events in perception.

Figure 4 shows the syllable-initial /b/-/w/ boundaries for birds and humans for both of the experimental manipulations. Based on the data in Fig. 4, three points are clear. First, the variation in duration of the adjacent phoneme following the target produced a significant change in the identification of the initial segment (/b/-/w/) for both birds and humans (with birds experiencing relatively larger boundary shifts). Second, variation in the duration of the nonadjacent phoneme /s/ only reliably influenced the perception of the target contrast for birds (although the human data trend in the expected direction). Third, boundary shifts are similarly large for Experiments 1 and 2 for birds, while the shift differences are significantly different between experiments for humans (significant shift in the first experiment, no shift in the second).

In Experiment 1, humans exhibited boundary locations and shifts analogous to those previously observed for rate-varying speech contexts using stop-semivowel continua (e.g., Miller and Liberman, 1979; Miller *et al.*, 1997; Wayland *et al.*, 1994). As seen in Fig. 5, however, it is known that absolute boundary locations vary between studies. In the present study, boundary locations moved from 39 to 45 ms (6 ms shift) after a change from short to long syllable (i.e., vowel) duration. Boundary locations for long syllables of about 300 ms in duration have previously been shown to be as short as 35–40 ms by Pisoni *et al.* (1983) or as long as 52–56 ms by Dent *et al.* (1997), Godfrey and Millay (1981), and Sinnott *et al.* (1998). Miller and Liberman's (1979) boundary of about 47 ms, falling between the others, is similar to the 45 ms boundary obtained for humans in the present study. Despite this range of boundary measurements, the overall magnitude of the shift for humans (approximately 6 ms) was comparable to other shifts for stop-glide stimuli of similar lengths (see Fig. 5).

The findings from Experiment 1 also add to the array of results where nonhumans exhibit speech context effects. Dent *et al.* (1997) used similar /ba/-/wa/ stimuli and observed peaks in discrimination functions for budgerigars at 40 ms for the short 120-ms stimuli and 50 ms for the 320-ms stimuli. Budgerigars have also previously exhibited discrimination peak shifts for other speech sounds and for sinewave

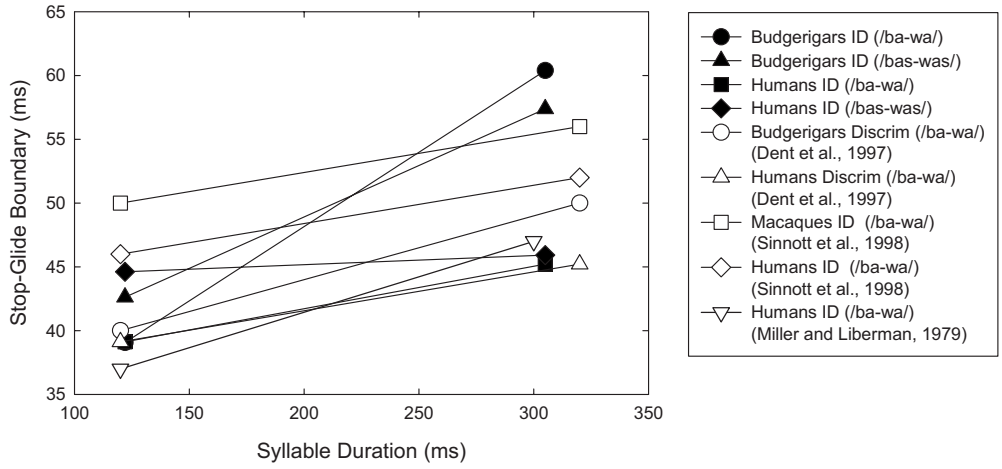


FIG. 5. Comparison of /b/-/w/ category boundaries as a function of syllable duration for the present study (black symbols) and several previous studies (white symbols).

patterns of speech in a manner similar to humans who were induced to hear the stimuli as speech (Best *et al.*, 1989; Dooling *et al.*, 1995). In Experiment 1 of the current study, mean boundary for the short syllables was 39 ms and moved to 60 ms for long syllables. This 21-ms boundary shift was much larger than the 6-ms shift observed in humans in this study and in previous studies (see above), although methodological differences and differing levels of experience with speech may be contributing to these boundary differences. Clearly though, the present results cannot be accounted for by suggesting that knowledge or experience with speech sounds in the natural communicative context is required to produce rate normalization effects.

Temporal limitations over which rate normalization effects may take place have not been tested previously in animals. For budgerigars in Experiment 2, the percept of a syllable initial target sound changed (for intermediate stimuli in the continua) after the addition of further-removed context (the final /s/ fricative) extending out to 340 ms in the most extreme case. It seems possible, for highly vocal birds that are relatively unfamiliar with human speech at least, that the temporal window for processing acoustic targets extends to over 350 ms and that this later-occurring information is readily used in the normalization process. Since the duration of the final /s/ altered classification of the initial consonant in birds, the /s/ appears to have been combined perceptually with the initial CV. Thus, the data for birds are comparable to previous human data showing an influence of a nonadjacent fricative (see Introduction). Therefore, these findings lend support to the notion that similar perceptual mechanisms such as durational contrast may be involved for both species.

When considering these results in terms of basic grouping principles, the effect for /s/ in birds means that temporal proximity is sufficient to cause the /s/ fricative to bind to the /ba/-/wa/ base, despite differences in periodicity (similarity).[1] A special speech mode of processing (e.g., Mattingly and Liberman, 1990; Remez *et al.*, 1994) was not available to the birds. By extension, this mode may not be necessary to explain rate normalization in humans. For example, the normalization or "contrastive" effects observed here may be a more general auditory phenomenon where birds are calibrating events based on relative duration, as humans often do. Principles like phonetic coherence (Remez *et al.*, 1994) may not be needed to explain findings of grouping speech or speechlike elements in studies of speaking rate normalization and duration contrast.

Since past studies with humans *have* found an influence of nonadjacent segments on boundary locations (e.g., Miller and Liberman, 1979; Newman and Sawusch, 1996), it remains possible that additional factors are involved in these effects. In this experiment, segment duration variation occurred after the initial /b/-/w/ transition *and* after a fixed 40 ms /a/ vowel. In the /bas/-/was/ experiments, perceptual processing of the /a/ vowel must precede that of the final /s/. It is possible that the short middle /a/ vowel was exerting too much of an influence on the humans for any influence of the later /s/ to emerge. For instance, it is possible that the short vowel /a/ was actually being used as an indicator for a fast rate (e.g., of speaking), thereby effectively reducing or eliminating any possible effect for the later /s/. The vowel is both closer and more spectrally similar to the to-be-judged target, so it also has the grouping principles of both proximity (temporal) and similarity acting to bind it to the /b/-/w/ base. The syllable-final /s/ has only proximity (although less so than the /a/) acting to group it to the initial contrast. Therefore, despite the fact that the stimuli were designed to fit within known temporal limits for processing synthetic speech, for humans with extensive experience with speech, it remains possible that the adjacent vowel between the target and the final /s/ exerted a strong enough effect that it overwhelmed the influence of the later-occurring and dissimilar sounding /s/ (see also Summerfield, 1981).

Since the identification methods used to test birds and humans were slightly different in the current experiments (e.g., in terms of stimulus presentation and feedback), it remains possible that these dissimilarities could account for some of the observed species differences. Despite all of these possible explanations, it is presently unclear why birds more readily used this later-occurring frication information in rate normalization compared to humans, who sometimes do and do not show normalization effects from later-occurring, nonadjacent segments.

V. CONCLUSIONS

The current findings from these animal participants add to the array of findings where nonhumans exhibit speech context effects in speech perception tasks. Here, rate normalization by budgerigars in an identification task extends previous speech discrimination results in birds. Further research may reveal analogous and general auditory mechanisms that may be operating in both birds and humans to account for the current rate normalization results. The process of normalizing for the rate at which events occur in perception seems to be a complex process but one that may not necessarily *require* phonetic mechanisms. Clearly, humans have access to and exploit information that can be used for phonetic coherence, and it seems plausible that phonetic coherence may operate along with Gestalt principles of organization in humans. The results from the birds clearly indicate that contiguity may be sufficient for causing these acoustic elements to cohere and suggests that for many cases of speech and speechlike sounds, only Gestalt-based principles of perceptual grouping may be necessary for this binding. The results reported here are consistent with the notion that rate normalization effects are widespread and not easily explained by any one principle of grouping.

ACKNOWLEDGMENTS

We would like to thank Jarrod Cone and numerous UB graduate and undergraduate research assistants for their invaluable assistance. This work was supported by grants from the DRF and NOHR.

[1]In addition to temporal proximity between the vowel and final fricative and the dissimilarity in source (a periodic vowel and an aperiodic fricative), there are also spectral similarities and dissimilarities between the vowel and the fricative. The fricative was synthesized with energy (noise) in F4 and F5. The frequencies of F4 and F5 in the vowel and the fricative

were the same. Consequently, there is some continuity in the presence of energy in these two frequency regions between the vowel and the fricative. However, the vowel has most of its energy in the lower formants (F1–F3) and there was no energy in these lower formants in the fricative. Thus, there were also substantial differences between the vowel and the fricative in the presence of energy in different frequency bands (formants). Overall, these continuities and discontinuities in the energy distribution could promote both a single unitary auditory stream and the formation of separate streams. The present description of the competing elements of temporal contiguity and source discontinuity is a simplification that neglects the competing influences of spectral continuity and dissimilarity.

Ainsworth, W. A. (**1973**). "Durational cues in the perception of certain consonants," Proc. Brit. Acoust. Soc. **2**, 1–4.

Best, C. T., Studdert-Kennedy, M., Manuel, S., and Rubin-Spitz, J. (**1989**). "Discovering phonetic coherence in acoustic patterns," Percept. Psychophys. **45**, 237–250.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound*, MIT, Cambridge, MA.

Byrd, D. (**1992**). "Preliminary results on speaker-dependent variation in the TIMIT database," J. Acoust. Soc. Am. **92**, 593–596.

Crystal, T. H., and House, A. S. (**1982**). "Segmental duration of American English vowels: An overview," J. Acoust. Soc. Am. **72**, 705–716.

Crystal, T. H., and House, A. S. (**1988**). "A note on the durations of fricatives in American English," J. Acoust. Soc. Am. **84**, 1932–1935.

Crystal, T. H., and House, A. S. (**1990**). "Articulation rate and the duration of syllables and stress groups in connected speech," J. Acoust. Soc. Am. **88**, 101–112.

Dent, M. L., Brittan-Powell, E. F., Dooling, R. J., and Pierce, A. (**1997**). "Perception of synthetic /ba/-/wa/ speech continuum by budgerigars (*Melopsittacus undulatus*)," J. Acoust. Soc. Am. **102**, 1891–1897.

Diehl, R. L., and Walsh, M. A. (**1989**). "An auditory basis for the stimuluslengtheffect in the perception of stops and glides," J. Acoust. Soc. Am. **85**, 2154–2164.

Dooling, R. J., Best, C. T., and Brown, S. D. (**1995**). "Discrimination of synthetic full-formant and sinewave /ra-la/ continua by budgerigars (Melopsittacus undulates) and zebra finches (*Taeniopygia guttata*)," J. Acoust. Soc. Am. **97**, 1839–1846.

Dooling, R. J., Lohr, B., and Dent, M. L. (**2000**). "Comparative hearing: Birds and reptiles," in *Hearing in Birds and Reptiles*, edited by R. J. Dooling, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 1–12.

Eimas, P., and Miller, J. (**1980**). "Contextual effects in infant speech perception," Science **209**, 1140–1141.

Fowler, C. A. (**1990**). "Sound-producing sources as objects of perception: Rate normalization and nonspeech perception," J. Acoust. Soc. Am. **88**, 1236–1249.

Godfrey, J., and Millay, K. (**1981**). "Discrimination of the 'tempo frequency change' cue," J. Acoust. Soc. Am. **69**, 1446–1448.

Jusczyk, P., Pisoni, D., Reed, M., Fernald, A., and Myers, M. (**1983**). "Infants' discrimination of the duration of a rapid spectrum change in nonspeech signals," Science **222**, 175–177.

Kidd, G. R. (**1989**). "Articulatory-rate context effects in phoneme identification," J. Exp. Psychol. Hum. Percept. Perform. **15**, 736–748.

Klatt, D. H. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–995.

Lotto, A. J., and Kluender, K. R. (**1997**). "Perceptual compensation for coarticulation by Japanese quail (*Coturnix japonica*)," J. Acoust. Soc. Am. **102**, 1134–1140.

Mattingly, I. G., and Liberman, A. M. (**1990**). "Speech and other auditory modules," in *Signal and Sense: Logical and Global Order in Perceptual Maps*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 501–520.

Miller, J. L. (**1981**). "Effects of speaking rate on segmental distinctions," in *Perspectives on the Study of Speech*, edited by P. D. Eimas and J. L. Miller (Eribaum, Hillsdale, NJ), pp. 39–74.

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., Grosjean, F., and Lomanto, C. (**1984**). "Articulation rate and its variability in spontaneous speech: A reanalysis and some implications," Phonetica **41**, 215–225.

Miller, J. L., and Liberman, A. M. (**1979**). "Some effects of later-occurring information on the perception of stop consonant and semivowel," Percept. Psychophys. **25**, 457–465.

Miller, J. L., O'Rourke, T. B., and Volaitis, L. E. (**1997**). "Internal structure of phonetic categories: Effects of speaking rate," Phonetica **54**, 121–137.

Minifie, F., Kuhl, P., and Stecher, B. (**1977**). "Categorical perception of [b] and [w] during changes in rate utterance," J. Acoust. Soc. Am. **62**, S79.

Newman, R. S., and Sawusch, J. R. (**1996**). "Perceptual normalization for speaking rate: Effects of temporal distance," Percept. Psychophys. **58**, 540–560.

Pisoni, D. B., Carrell, T. D., and Gans, S. J. (**1983**). "Perception of the duration of rapid spectrum changes in speech and nonspeech signals," Percept. Psychophys. **34**, 314–322.

Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., and Lang, J. K. (**1994**). Psychol. Rev. **101**, 129–156.

Sawusch, J. R., and Newman, R. S. (**2000**). "Perceptual normalization for speaking rate II: Effects of signal discontinuities," Percept. Psychophys. **62**, 285–300.

Sinnott, J. M., Brown, C. H., and Borneman, M. A. (**1998**). "Effects of syllable duration on stop-glide identification in syllable initial and syllable-final position by humans and monkeys," Percept. Psychophys. **60**, 1032–1043.

Summerfield, Q. (**1981**). "Articulatory rate and perceptual constancy in phonetic perception," J. Exp. Psychol. Hum. Percept. Perform. **7**, 1074–1095.

Wade, T. and Holt, L. L. (**2005**). "Effects of later-occurring non-linguistic sounds on speech categorization," J. Acoust. Soc. Am. **118**, 1701–1710.

Wayland, S. C., Miller, J. L., and Volaitis, L. E. (**1994**). "The influence of sentential speaking rate and the internal structure of phonetic categories," J. Acoust. Soc. Am. **95**, 2694–2701.

Behavioral measures of signal recognition thresholds in frogs in the presence and absence of chorus-shaped noise

Mark A. Bee[a)]
Department of Ecology, Evolution, and Behavior, University of Minnesota, 100 Ecology, 1987 Upper Buford Circle, St. Paul, Minnesota 55108

Joshua J. Schwartz
Department of Biology and Health Sciences, Pace University, Pleasantville, New York 10570

(Received 4 June 2009; revised 11 August 2009; accepted 12 August 2009)

Anuran amphibians are superb animal models for investigating the mechanisms underlying acoustic signal perception amid high levels of background noise generated by large social aggregations of vocalizing individuals. Yet there are not well-established methods for quantifying a number of key measures of auditory perception in frogs, in part, because frogs are notoriously difficult subjects for traditional psychoacoustic experiments based on classical or operant conditioning. A common experimental approach for studying frog hearing and acoustic communication involves behavioral phonotaxis experiments, in which patterns of movement directed toward sound sources indicate the subjects' perceptual experiences. In this study, three different phonotaxis experiments were conducted using the same target signals and noise maskers to compare different experimental methods and analytical tools for deriving estimates of signal recognition thresholds in the presence or absence of "chorus-shaped noise" (i.e., artificial noise with a spectrum similar to that of real breeding choruses). Estimates of recognition thresholds based on measures of angular orientation, response probabilities, and response latencies were quite similar in both two-choice and no-choice phonotaxis tests. These results establish important baselines for comparing different methods of estimating signal recognition thresholds in frogs tested in various masking noise conditions.
 [DOI: 10.1121/1.3224707]

PACS number(s): 43.80.Nd, 43.80.Lb [JAS] Pages: 2788–2801

I. INTRODUCTION

Animals that signal acoustically in large social aggregations, such as choruses (e.g., Gerhardt and Huber, 2002) and communal crèches (e.g., Aubin and Jouventin, 2002), represent ideal model systems for investigating how animals cope with the problems of noise (Schwartz and Freeberg, 2008). For such species, the potential impacts of auditory masking and interference should be especially severe because of the high degree of overlap among the spectral and temporal properties of the concurrent signals of other conspecific individuals. In humans, these impacts give rise to the so-called "cocktail-party problem," which refers to our difficulty perceiving speech in noisy social settings (Cherry, 1953; Bee and Micheyl, 2008). Anuran amphibians (frogs and toads) represent one taxonomic group of non-human animals for which cocktail-party-like problems are likely to be pronounced in acoustic communication (Narins and Zelick, 1988; Feng and Schul, 2007). In many anurans, reproduction takes place in large and often dense breeding aggregations in which males produce loud advertisement calls that are both necessary and sufficient to attract gravid females (Gerhardt and Huber, 2002).

Here, we report results from the latest in a series of studies aimed at understanding how females of Cope's gray treefrog (*Hyla chrysoscelis*) perceive the acoustic signals of sexually advertising males amid the high levels of noise present in breeding choruses (see Bee, 2007, 2008a, 2008b; Bee and Swanson, 2007; Swanson *et al.*, 2007). The primary goal of this study was to evaluate various behavioral and analytical methods for estimating "signal recognition thresholds" in the presence and absence of "chorus-shaped noise" using phonotaxis experiments. We used both no-choice and two-choice phonotaxis tests (Gerhardt, 1995; Ryan and Rand, 2001) to measure as a function of signal level the responses of females to a conspecific advertisement call presented in the presence or absence of noise, and we explored a range of threshold criteria for estimating signal recognition thresholds. Our study was aimed at evaluating procedures that could allow researchers to use phonotaxis experiments as a tool to investigate more systematically a number of empirical questions concerning how frogs communicate in noisy social environments (reviewed in Bee and Micheyl, 2008). Our approach is one inspired by studies of human speech perception in noise, which commonly measure the "speech reception threshold" (SRT) (Plomp, 1978; Plomp and Mimpen, 1979a, 1979b). The SRT corresponds to the sound level at which a segment of speech must be presented to listeners in order for intelligibility to reach some predetermined threshold level as measured by the percentage of target words correctly repeated. Framed more broadly, the SRT relies on a correct response from a listener in the form of a species-typical behavior that is used to measure thresholds

[a)]Author to whom correspondence should be addressed. Electronic mail: mbee@umn.edu

for recognizing conspecific vocal signals. In this study, we operationally defined signal recognition as occurring when a female exhibited a correct response (phonotaxis) with respect to a conspecific signal (see discussion of terminology in Gerhardt and Huber, 2002 and Sec. VI A.); we operationally defined the signal recognition threshold as the minimum signal level required to elicit phonotaxis behavior exceeding a pre-determined criterion level of correct responses.

II. METHODS

A. The study system

Gray treefrogs represent a cryptic species complex comprising Cope's gray treefrog (*H. chrysoscelis*, a diploid) and the eastern gray treefrog (*H. versicolor*, a tetraploid) (Holloway *et al.*, 2006). Both species range widely throughout much of eastern North America and have been the subjects of extensive and detailed investigations of hearing and acoustic communication, species recognition, reproductive behavior, and sexual selection (reviewed in Gerhardt, 2001). Like many other frogs, male gray treefrogs form spring breeding choruses in which they produce advertisement calls with amplitudes ranging from about 85 to 93 dB sound pressure level (SPL) rms (96–104 dB peak) measured at a distance of 1 m (Gerhardt, 1975). The nearly continuous background noise levels in dense gray treefrog choruses can be quite intense, reaching sustained levels of 70–80 dB SPL (Schwartz *et al.*, 2001; Swanson *et al.*, 2007; Vélez and Bee, unpublished data). Gray treefrog advertisement calls comprise a series of discrete pulses produced at species-specific rates; in *H. chrysoscelis*, the pulse rate (about 40–50 pulses/s) is an important species recognition cue (Gerhardt, 2001; Schul and Bush, 2002). The call has a bimodal frequency spectrum, with a fundamental frequency in *H. chrysoscelis* of about 1200–1400 Hz that has an amplitude of −5 to −10 dB relative to the dominant frequency of about 2400–2800 Hz (Gerhardt, 2001). The spectrum of the background noise generated in gray treefrog choruses is similar to that of the advertisement call (Swanson *et al.*, 2007).

B. Subjects

We collected 132 pairs of *H. chrysoscelis* in amplexus between 2100 and 0100 h during the 2008 breeding season (May and June) from wetlands in the Carver Park Reserve (Carver Co., Chaska, MN). Pairs were returned to the laboratory and kept at 2 °C to delay egg deposition until the females were tested (usually within 24 h). Pairs were returned to their original location of capture after testing. Additional details about collections and handling of frogs have been published elsewhere (Bee, 2007, 2008a, 2008b; Bee and Swanson, 2007). Of the 132 females collected and tested for this study, 120 females met an inclusion criterion that required them to finish a designated series of tests to be included as subjects in statistical analyses.

C. General testing procedures

On the day of testing, we transferred pairs to a 20 °C incubator where they remained at least 1 h prior to testing until their body temperature reached 20 °C (±1 °C). Phonotaxis experiments were conducted in two temperature-controlled, hemi-anechoic sound chambers [Industrial Acoustics Co., (IAC), Bronx, NY; inside dimensions $L \times W \times H$: $300 \times 280 \times 216$ cm^3 and $220 \times 280 \times 216$ cm^3]. The inside walls and ceiling of the chambers were painted dark gray and treated with IAC's Planarchoic™ treatment to reduce reverberation. The chambers had vibration-isolation floors that were covered in dark gray carpet. We controlled the temperature inside the chambers at 20 °C (±2 °C), which is a typical temperature at which gray treefrogs breed. With their ventilation units running, the SPL of each chamber's ambient noise floor ranged between 2 and 12 dB SPL (fast rms, flat weighting) in the 1/3-octave bands between 500 and 4000 Hz, which spans the frequency range of interest in this study. The frequency responses of the playback systems in the two chambers were flat (±3 dB) over the same frequency range.

We used ADOBE AUDITION v1.5 (Adobe Systems Inc., San Jose, CA) to broadcast digital acoustic stimuli (20 kHz sampling rate, 16-bit resolution) from a Dell Computer (Optiplex GX620 or GX745; Dell Computer Corp., Round Rock, TX) located outside each chamber. Each computer was interfaced with an M-Audio FireWire 410 multichannel soundcard (M-Audio USA, Irwindale, CA), and the output of the soundcard was amplified using either a Sonamp 1230 (Sonance, San Clemente, CA) or HTD 1235 (Home Theater Direct Inc., Plano, TX) multichannel amplifier.

All behavioral tests were conducted under infrared (IR) illumination provided by two IR light sources (Noldus Information Technology Inc., Leesburg, VA) in each sound chamber that were mounted near the ceilings on opposite walls. We monitored behavioral responses in real time from outside the chamber and recorded them direct to digital video using real-time MPEG encoders (MVR1000SX or MPEGPRO EMR, Canopus, San Jose, CA) interfaced with an overhead, IR-sensitive Panasonic WV-BP334 video camera (Panasonic Corporation of North America, Secaucus, NJ) mounted from the center of each sound chamber's ceiling.

Phonotaxis tests were performed in 2-m diameter circular test arenas (one per sound chamber) with walls that were 60-cm high and constructed from acoustically transparent but visually opaque black cloth and hardware cloth. The sound chambers' carpeted floors served as the test arena floors, which were divided into 24 15° arcs along their perimeters. Synthetic advertisement calls (see below) were broadcast through A/D/S L210 speakers (Directed Electronics, Inc., Vista, CA) that were placed in the center of the 15° arcs on the floor just outside the walls of the test arenas and directed toward the arenas' centers. The positions of speakers around an arena's perimeter were varied on a regular basis to eliminate any possibility of a directional response bias in our sound chambers. Masking noise (see below) was broadcast through a Kenwood KFC-1680ie speaker (Kenwood USA Corporation, Long Beach, CA) suspended from the ceiling of each chamber 190 cm above the center of the test arena. The overhead speaker created a uniform (±1.5 dB) noise level across the entire floor of an arena.

At the beginning of each phonotaxis test, a female was placed in a holding cage located on the floor of the sound chamber at the center of the test arena. The holding cage consisted of a shallow, acoustically transparent cup (9-cm diameter; 2-cm height) with a lid that could be removed using a rope and pulley system operated from outside the sound chamber to allow females unrestricted movement within the test arena. Stimulus broadcasts began after a 1.5-min acclimation period and were continued throughout the duration of a test. Females were remotely released from the holding cage after 30 s of signal presentation. Sound levels were measured and calibrated prior to testing by placing the microphone of a Larson-Davis System 834 (Larsen Davis, Depew, NY) or a Brüel & Kjær type 2250 (Brüel & Kjær Sound & Vibration Measurement A/S, Nærum, Denmark) sound level meter at the approximate position of a subject's head at the central release point. Below, we report all signal and noise levels and threshold estimates in units of dB SPL (re 20 μPA, fast rms, *C*-weighted).

III. EXPERIMENT 1: POPULATION-LEVEL RECOGNITION THRESHOLDS IN TWO-CHOICE TESTS

In this experiment, we explored a number of methods to estimate recognition thresholds based on pooling data from a group of subjects tested using a traditional two-choice experimental design (Gerhardt, 1995). Each subject was presented with a target signal (conspecific call) that alternated in time with a non-target signal (heterospecific call). The rms SPLs of the two signals were always equal in each test but were varied systematically between different tests. Subjects experienced each signal level only one time, and subjects were assigned randomly to one of two noise conditions, a "no-noise group" (N=20) or a "noise group" (N=20). Our estimates of recognition thresholds for each group are based on performance measures that we derived from the behavior of the entire pool of subjects in that group; therefore, we regard these estimates as being "population level" in the sense that we did not attempt to estimate a threshold for each individual as is more common in traditional psychoacoustic experiments (Klump *et al.*, 1995).

A. Methods

1. Experimental design

Experiment 1 was based on a 9 signal level (within subjects) × 2 noise condition (between subjects) factorial design. Across nine "treatment conditions," we varied the SPLs of the target and non-target signals across nine nominal levels ranging from 37 to 85 dB SPL in 6-dB steps. The SPLs of both the target and the non-target signal were adjusted using software control of the M-Audio soundcard to have the nominal signal level at the position of a subject's head at the central release site located 1 m from the speakers. The source level of the signals remained constant during a test, and the order of different signal levels across the nine treatment conditions was determined randomly for each subject.

The acoustic signals were synthetic advertisement calls that were created using custom-made software and modeled after natural calls having spectral and temporal properties close to the averages (corrected to 20 °C) of calls recorded in local Minnesota populations (Bee, unpublished data). The target signal [Fig. 1(a)] was a conspecific (*H. chrysoscelis*) call that was 693 ms in duration and consisted of 32 pulses (11-ms pulse duration) delivered at a rate of 45.5 pulses/s (22-ms pulse period). The non-target (heterospecific) signal was an advertisement call of the closely related eastern gray treefrog (*H. versicolor*), was 690 ms in duration, and consisted of 12 pulses (30-ms pulse duration) delivered at a rate of 16.7 pulses/s (60-ms pulse period). In both signals, each pulse consisted of two harmonically related, phase-locked sinusoids with frequencies of 1.25 and 2.5 kHz and relative amplitudes of −6 and 0 dB, respectively [Fig. 1(b)]. Both signals repeated with a period of 5 s, which is within the range of call periods in local populations (corrected to 20 °C), and alternated so that equal durations of silence preceded and followed each signal. For half of the subjects tested, the alternating stimulus sequence began with the target signal; the other half of subjects heard the non-target signal first. Where possible, we included this "call order"

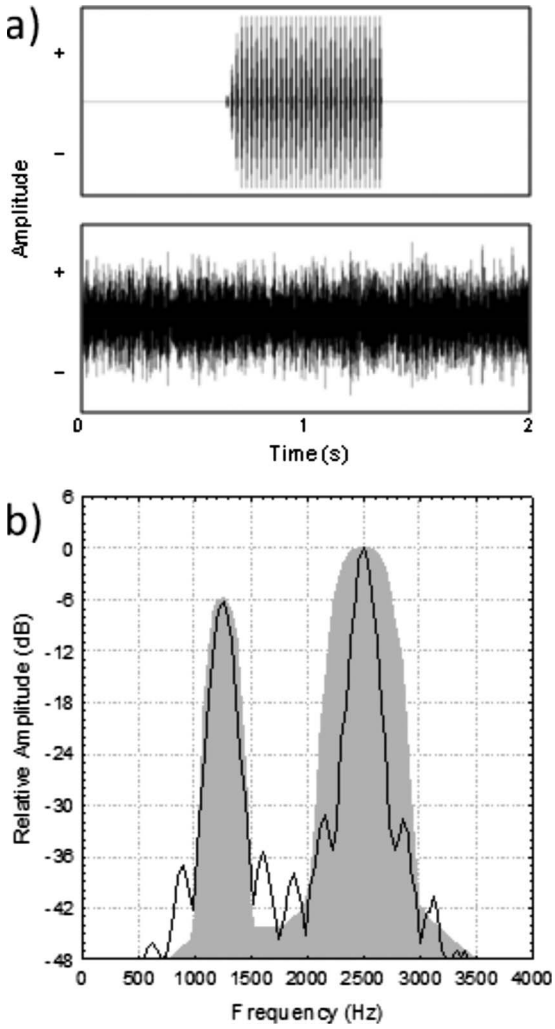


FIG. 1. Target signal and chorus-shaped noise. (a) Waveforms of the synthetic *H. chrysoscelis* call comprising the target signal (top) and a segment of chorus-shaped noise (bottom). (b) Power spectra showing the spectral profile of the target signal (black line) relative to that of the chorus-shaped noise (gray area).

effect in our statistical analyses. The target and non-target signals were broadcast from separate speakers located directly opposite each other around the circular test arena (i.e., 2 m and 180° apart).

All subjects were tested individually in a sequence of 13 phonotaxis tests and were given a 5- to 10-min timeout period inside the incubator between consecutive tests. A test sequence began with a "reference condition" and then alternated between three treatment conditions and another reference condition until all nine treatment conditions had been tested. Each sequence ended with a final test of the reference condition. Hence, each subject was tested a total of 13 times (9 treatment conditions + 4 reference conditions). For all subjects, the reference condition consisted of broadcasting the alternating target and non-target signals at 85 dB SPL in the absence of any broadcast masking noise.

Subjects in the no-noise group always experienced the signals without the broadcast of any additional masking noise. For subjects assigned to the noise group, we broadcast the signals in the presence of a chorus-shaped noise that had a long-term spectrum with acoustic energy at audio frequencies characteristic of gray treefrog choruses (Fig. 1). We used MATLAB v7.6.0 to create five different exemplars of chorus-shaped noise in the following way. For each exemplar, we first filtered two copies of the same 6-min long white noise to create two narrowband noises centered at 1250 and 2500 Hz. The low-frequency band was created using a band-pass finite impulse response (FIR) filter of order 300, pass-band frequencies of 1200 and 1300 Hz, and stop-band frequencies of 1000 and 1500 Hz. The high-frequency bands were created using a band-pass FIR filter of order 150, with pass-band frequencies of 2400 and 2600 Hz and stop-band frequencies of 2000 and 3000 Hz. Both FIR filters had pass-band ripples of 0.1 Hz and stop-band attenuations of 60 dB. The peak amplitude of the low-frequency band was attenuated 6 dB relative to that of the high-frequency band and then both noises were digitally added to create a single chorus-shaped noise (Fig. 1). During behavioral tests with subjects in the noise group, we began broadcasts of the noise 30 s prior to the onset of the alternating signals and the broadcast continued over the duration of the test. The sound level of each noise exemplar was calibrated to be 70 dB SPL (LC_{eq}) at the approximate position of a subject's head at the release site at the center of the test arena.

2. Data analysis

We scored a "correct response" if the subject touched the arena wall inside the 15° arc in front of the speaker that was broadcasting the target (conspecific) signal within 5 min of being released. In essence, a correct response is one that would likely result in the selection of a conspecific mate under natural conditions, and hence demonstrates correct recognition of a conspecific vocalization. In keeping with our operational definition of recognition (Sec. I), we scored an "incorrect response" if either of the following two conditions were met: (i) a subject touched the arena wall in the 15° arc in front of the speaker broadcasting the non-target (heterospecific) signal; (ii) the subject failed to touch the wall in front of either speaker within 5 min. These two behavioral outcomes might have different underlying causes (e.g., failed recognition versus failed detection, respectively); however, both outcomes are inconsistent with our operational definition of recognition. Although phonotaxis toward the non-target (heterospecific) signal indicates acceptance of the signal as that of an appropriate mate (e.g., Ryan and Rand, 1993), we consider it an incorrect response because such a choice in nature would result in the production of offspring that were inviable, infertile, or of reduced attractiveness (Gerhardt *et al.*, 1994). An incorrect response was only considered legitimate if the subject exhibited a correct response in the next reference condition in sequence. Only subjects that exhibited correct responses in all four reference conditions were included in statistical analyses. We also excluded from the final data set any subject that required more than twice as long to respond in the final reference condition compared with the first reference condition. Such procedures ensure the validity of no responses (or slow responses) in treatment conditions by confirming that subjects remain highly motivated to respond over the duration of the test sequence (Bush *et al.*, 2002; Schul and Bush, 2002). Of the 46 females tested in this experiment, 6 did not meet our response criteria, yielding a final sample size of $N=40$. None of the subjects in this experiment had been tested previously. We used analysis of variance (ANOVA) to assess differences in response latency across the four reference conditions.

We assessed differences in three related response variables as functions of noise condition and signal level.

(a) *Angular orientation.* We assessed the directedness of phonotaxis toward the target signal by measuring a subject's angular orientation relative to the position of the target playback speaker (designated as 0°) when it first reached a distance of 20 cm away from the central release point. We chose a distance of 20 cm as a compromise between analyzing the angles at which females exited the release cage and the angles at which they first touched the arena wall 1 m away. To exit our release cage, females had to climb over a 2-cm high barrier. In our experience, females sometimes climb over this barrier in one direction only to quickly reorient and initiate movement in a different direction while still located immediately adjacent to the release cage. Thus, we believe allowing the females to freely move about on the arena floor (outside of the release cage) is a relatively more accurate measure of orientation behavior. However, we also believe that restricting the measurement distance to 20 cm minimizes any cues related to the variation in signal levels experienced by moving about in the sound field on the arena floor. According to both the inverse square law and our own empirical measurements in the sound chambers, the gain in signal level experienced by moving 20 cm closer to a source originally located 1 m away is less than 2 dB, which is much less than the 6-dB step-size we used between adjacent signal levels. We used V-tests (Zar, 1999) to test the null hypothesis that angles at 20 cm were uniformly distributed against the alternative hypothesis that subjects oriented toward the target signal at 0°. We estimated an upper threshold

bound as the lowest signal level at which statistically significant orientation occurred at that level and also at all higher levels; the next lowest signal level was used as an estimate of a lower bound (LB). We then computed a recognition threshold as the average of the upper bound (UB) and LB using the following equation:

$$\text{recognition threshold} = 10 \log_{10}\left(\frac{10^{(\text{UB}/10)} + 10^{(\text{LB}/10)}}{2}\right). \quad (1)$$

(b) *Response probabilities.* We used generalized linear models (proc GENMOD in SAS) to examine differences in the probability of a correct response (1=correct response, 0=incorrect response) as functions of noise condition, signal level, and call order. These models are based on the binomial distribution, the logit link function, and the use of generalized estimating equations (GEEs) for estimating within-subjects effects (Horton and Lipsitz, 1999). We explored a range of threshold criteria based on extrapolated and interpolated values along the best-fit logistic regression curves relating response probability to signal level.

We also used data on raw response probabilities to estimate thresholds in two additional ways. First, following Beckers and Schul (2004), we estimated the UB of a recognition threshold as the lowest signal level at which greater than 50% of subjects exhibited correct responses at that level and also at all higher levels. LBs were determined as the next lowest signal level, and the recognition threshold was determined using Eq. (1). Second, we estimated thresholds based on determining as an UB the lowest signal level at which the proportion of females touching the arena wall in front of the target speaker exceeded the chance probability of doing so at that signal level and also at all higher levels. In a series of preliminary experiments, we estimated the false alarm rate of our response criterion by examining the behavior of female gray treefrogs in the test arena when no signals were presented in the presence or absence of chorus-shaped noise. In assessing the proportion of subjects that touched the arena wall in a 15° arc in front of a silent speaker within 5 min, we found that approximately 10%–20% would be expected to do so by chance in the absence of any signal using our protocol (Vélez and Bee, unpublished data). Therefore, we used Eq. (1) to estimate a threshold based on taking as an UB the lowest signal level at which the proportion of subjects exhibiting correct responses was significantly greater than 0.20 (two-tailed binomial tests) at that level and also at all higher levels; the next lowest signal level was taken as the LB.

(c) *Phonotaxis scores.* As a third and final measure of behavioral responsiveness, we calculated "phonotaxis scores," which normalize the latency in a treatment condition by dividing the average latency from the two most temporally proximal reference conditions by the latency in the treatment condition (Bush *et al.*, 2002; Schul and Bush, 2002; Beckers and Schul, 2004). A phonotaxis score of 1.0 thus indicates that the latency in the treatment condition equals that in the temporally proximal reference conditions; scores greater than 1.0 and less than 1.0 indicate latencies that are shorter and longer, respectively, than those in the reference conditions. We assigned a score of 0.0 when subjects exhibited incorrect responses. We analyzed phonotaxis scores using a 9 signal level (within subjects) × 2 noise condition (between subjects) × 2 call order (between subjects) ANOVA and report the Greenhouse and Geisser (1959) corrected P values for tests involving within-subjects effects with more than a single numerator degree of freedom. We used curve fitting procedures to separately compute the best-fit sigmoid function relating mean phonotaxis scores to signal level in the two noise conditions according to the following equation:

$$\text{phonotaxis score} = a/(1.0 + e^{(-(\text{signal level}-b)/c)}) + d, \quad (2)$$

where a, b, c, and d are the fitted parameters that minimized the sum of the squared absolute error, and e is the base of the natural logarithm. We chose to fit our data with sigmoid curves because such curves often characterize the shapes of both psychometric functions generated in psychophysical experiments and neuronal rate-level functions generated in electrophysiological studies. We used the fitted sigmoid equations to explore a range of criteria for estimating recognition thresholds from phonotaxis scores.

B. Results and discussion

Subjects remained highly motivated to respond over the duration of the test sequence as evidenced by their uniformly strong orientation toward the target signal in the four reference conditions (Table I). Averaged across all four reference conditions and both noise conditions (i.e., the noise and no-noise groups), subjects made their correct responses with a mean (±SD) latency of 76.3 ± 28.6 s. There were no significant differences in latency across the four reference conditions ($F_{3,108}=1.9$, $P=0.1461$). There were also no significant differences in latency between the two noise conditions ($F_{1,36}=4.0$, $P=0.0523$) or according to which signal (target or non-target) was broadcast first ($F_{1,36}=3.5$, $P=0.0695$), nor were there any significant interactions between any of the main effects ($0.1461 < Ps < 0.9702$).

1. Angular orientation

Based on our measures of angular orientation at a distance of 20 cm, the difference in recognition thresholds between the no-noise and noise groups was 30 dB. In the no-noise group, subjects oriented significantly in the direction of the conspecific target signal at signal levels of 43 dB and higher (Table I). Using 43 dB as an UB and 37 dB as the LB, we calculated a recognition threshold of 41 dB in the no-noise group. In the presence of chorus-shaped noise, signal levels of 73 dB and higher elicited significant orientation toward the target signal (Table I). Taking 67 dB as a LB we calculated a recognition threshold of 71 dB in the noise group.

TABLE I. Results of circular statistical analyses for Experiment 1 (two-choice tests).

Noise condition	Signal condition	Mean vector ($\mu°$)	Length of mean vector (r)	Circular SD (deg)	N	V	P
No-noise	Reference 1	−5	0.89	28	20	0.88	<0.0001
	Reference 2	−6	0.90	27	20	0.89	<0.0001
	Reference 3	11	0.93	22	20	0.91	<0.0001
	Reference 4	0	0.84	34	20	0.84	<0.0001
	37 dB	9	0.24	97	16	0.23	0.0940
	43 dB	26	0.51	67	16	0.46	0.0040
	49 dB	−24	0.54	64	18	0.49	0.0010
	55 dB	6	0.54	64	18	0.53	0.0005
	61 dB	20	0.54	64	20	0.51	0.0005
	67 dB	−3	0.84	34	20	0.84	<0.0001
	73 dB	−1	0.78	41	20	0.78	<0.0001
	79 dB	−2	0.89	27	20	0.89	<0.0001
	85 dB	−2	0.95	19	20	0.95	<0.0001
Noise	Reference 1	5	0.98	12	20	0.98	<0.0001
	Reference 2	−5	0.96	17	20	0.96	<0.0001
	Reference 3	1	0.95	18	20	0.95	<0.0001
	Reference 4	2	0.96	16	20	0.96	<0.0001
	37 dB	−164	0.31	87	18	−0.30	0.9650
	43 dB	−175	0.15	113	17	−0.14	0.7980
	49 dB	60	0.31	88	19	0.16	0.1710
	55 dB	−14	0.24	97	18	0.23	0.0810
	61 dB	−75	0.36	82	18	0.09	0.2910
	67 dB	139	0.16	110	18	−0.12	0.7670
	73 dB	−22	0.46	72	20	0.43	0.0030
	79 dB	−2	0.97	14	19	0.97	<0.0001
	85 dB	−3	0.94	19	20	0.94	<0.0001

2. Response probabilities

The proportion of subjects that exhibited correct responses increased as a function of increasing signal level in both the no-noise and noise groups, and this level-dependent increase began at higher signal levels in the presence of chorus-shaped noise [Figs. 2(a)–2(c)]. In the generalized linear model for this experiment, the parameter relating response probability to signal level was significantly different from zero ($\chi^2=29.3$, $P<0.0001$, df=1). The parameters for noise condition ($\chi^2=3.4$, $P=0.0661$, df=1), call order ($\chi^2=0.1$, $P=0.7037$, df=1), and the interaction between signal level and noise condition ($\chi^2=0.03$, $P=0.8545$, df=1) were not different from zero. Subsequent contrast analyses of least-squares means, however, revealed significant differences between the no-noise and noise groups ($\chi^2=23.3$, $P<0.0001$, df=1). The proportions of correct responses in the no-noise group were significantly greater than zero ($\chi^2=25.7$, $P<0.0001$, df=1) while those in the noise group were not ($\chi^2=0.7$, $P=0.4021$, df=1). There was no significant effect according to which signal initiated the alternating broadcasts ($\chi^2=0.2$, $P=0.6934$, df=1). In Table II, we summarize threshold estimates based on different threshold criteria expressed as the probability of a correct response (p') along the fitted logistic regression functions [Fig. 2(c)]. Differences in thresholds between the no-noise and noise groups were consistently close to 20 dB regardless of the threshold criterion (Table II). There was no single threshold criterion along the fitted logistic regression curves that yielded absolute threshold estimates for both the no-noise and noise groups (Table II) that were simultaneously consistent with those based either on angular orientation (Sec. III B 1) or on the raw proportions of 0.50 or 0.20, which we describe next.

Estimates of recognition thresholds based on the raw proportions of females exhibiting correct responses were consistently 30 dB higher in the noise group compared with the no-noise group. The lowest signal levels at which at least 50% of subjects exhibited correct responses were 43 dB (11 of 20 responded) and 73 dB (19 of 20 responded) in the no-noise and noise groups, respectively. These UBs, and their corresponding LBs of 37 and 67 dB, yielded threshold estimates of 41 and 71 dB, respectively. The proportion of females that exhibited correct responses was significantly greater than the expected false alarm rate of 0.20 at signal levels of 67 dB and higher in the noise group, yielding a threshold estimate of 65 dB for this group. In the no-noise group, however, the proportion of females exhibiting correct responses significantly exceeded 0.20 at the lowest signal level of 37 dB and all higher levels. Therefore, to compute a threshold estimate for this group, we assumed that subjects would not have exhibited phonotaxis at the next lowest signal level in series, which would have been 31 dB (i.e., a 6-dB step down from the lowest signal level used). This assumption is reasonable given that Beckers and Schul (2004) reported that 4 of 11 (36.4%) and 5 of 11 (45.5%)

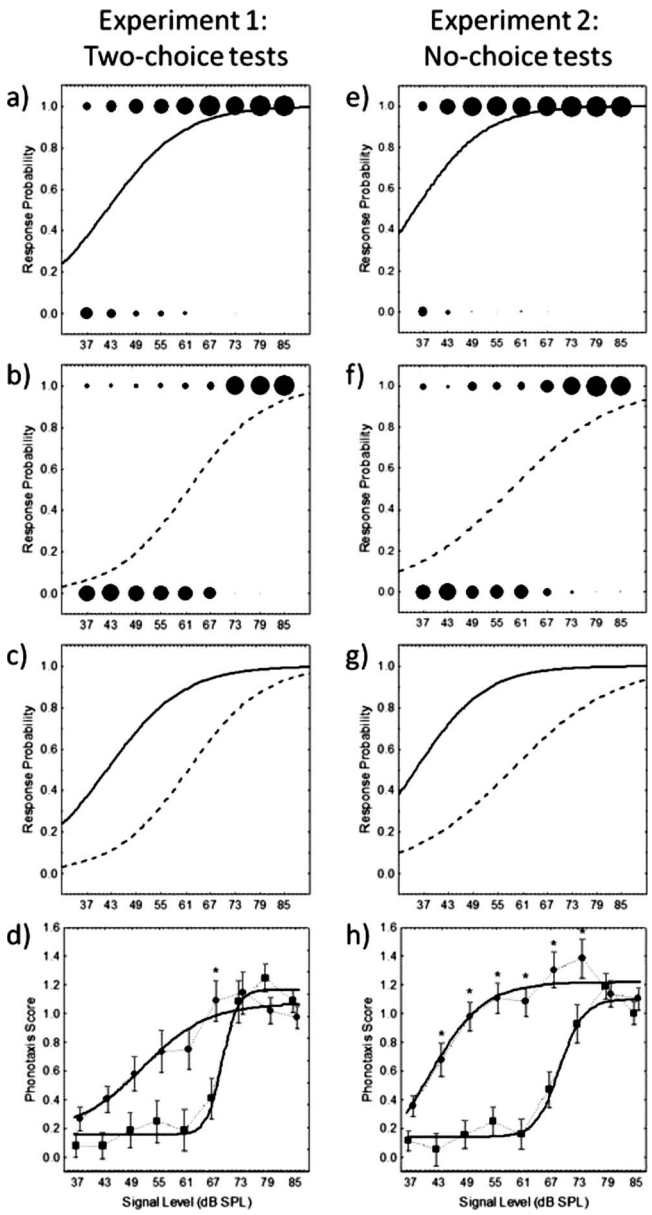


FIG. 2. Response probabilities and phonotaxis scores based on two-choice tests in Experiment 1 (left) and no-choice tests in Experiment 2 (right). [(a) and (e)] Response probabilities in the no-noise groups showing the numbers of individuals (total $N=20$) that exhibited the two types of responses, correct responses (1.0) and incorrect or no responses (0.0). For each response type, the relative sizes of the two paired points at each signal level depict the numbers of individuals exhibiting that response type, with the largest point corresponding to 20 individuals, and the smallest point corresponding to 1 individual (the absence of a point corresponds to zero individuals). The smooth curve represents the best-fit logistic regression function fitted to the response probabilities depicted in the figure. [(b) and (f)] Response probabilities and logistic regression functions (as in panels a and e) for the noise groups. [(c) and (g)] Comparison of the fitted logistic regression functions for the no-noise groups (solid lines) and the noise groups (dashed lines) for each experiment. [(d) and (h)] Mean ($\pm$s.e.m.) phonotaxis scores with best-fit sigmoid curves from Eq. (2) for responses in the no-noise groups (circles) and the noise groups (squares) for each experiment. $^{*}P<0.05$ in a Bonferroni *post hoc* test comparing phonotaxis scores in the no-noise and noise groups at each signal level.

female *Hyla versicolor* responded to a conspecific call in no-choice tests in the absence of masking noise at signal levels of 31 dB (two-tailed binomial test of $p \geq 0.20$; $P=0.2470$) and 37 dB (two-tailed binomial test of $p \geq 0.20$; $P=0.0504$). At a sample size equivalent to that used in the present study ($N=20$), the proportions reported by Beckers and Schul (2004) would be significantly greater than 0.20 at the 37-dB signal level but not at a level of 31 dB. Therefore, taking 31 and 37 dB as the LB and UB, we estimated a threshold of 35 dB for this group. We acknowledge that this estimate must be accepted with some caution because signal levels below 37 dB were not tested in the present study.

We performed one additional set of analyses based on examining the choices made by those females (out of 20) that touched the wall in front of the speaker broadcasting the target (conspecific) signal or the non-target (heterospecific) signal. In traditional analyses of results from two-choice phonotaxis tests with frogs, the former are typically regarded as a preferential "choice" of the conspecific call over the heterospecific call; the latter are regarded as a choice of the heterospecific call. Considering only those females that made a choice, there was a significant preference for the target (conspecific) signal at all signal levels tested in the no-noise group (Table III). By contrast, a signal level of at least 73 dB was necessary to elicit a significant preference in favor of the target (conspecific) signal in the presence of the chorus-shaped noise. If 37 and 73 dB are taken as UBs of a threshold estimate in the no-noise and noise groups, respectively, and corresponding values of 31 and 67 dB are taken as LBs, then recognition threshold estimates based on measures of behavioral discrimination between conspecific and heterospecific calls would be 35 and 71 dB in the no-noise and noise groups, respectively. The difference between these estimates of recognition threshold (36 dB) based on the responses of females exhibiting choices is therefore greater than those derived from measures of angular orientation (30 dB; Sec. III B 2) and other measures of response probability based on raw or fitted values ($\approx$20–30 dB).

TABLE II. Behavioral recognition thresholds in the no-noise and noise groups of Experiment 1 (two-choice tests) as functions of threshold criteria based on fitted response probabilities and phonotaxis scores.

Response variable	Threshold criterion	Estimated threshold signal levels (dB SPL)		Threshold difference (dB)
		No-noise	Noise	
Response probability (p')	0.2	29.2	49.2	20.0
	0.3	34.2	54.1	19.9
	0.4	38.2	58.0	19.8
	0.5	42.0	61.7	19.7
	0.6	45.7	65.3	19.6
	0.7	49.8	69.3	19.5
	0.8	54.8	74.2	19.4
	0.9	62.2	81.5	19.3
Phonotaxis scores (ps')	0.2	23.2	63.6	40.4
	0.3	37.9	65.8	27.9
	0.4	43.4	66.9	23.5
	0.5	47.2	67.7	20.5
	0.6	50.6	68.5	17.9
	0.7	53.8	69.1	15.3
	0.8	57.3	69.8	12.5
	0.9	61.7	70.6	8.9

TABLE III. Number (and percentages) of subjects choosing the target (conspecific) signal and the non-target (heterospecific) signal as a function of signal level in the presence or absence of chorus-shaped noise with results from two-tailed binomial tests of the null hypothesis that $p=0.50$.

Noise condition	Signal level	Target signal	Non-target signal	Total	P
No-noise	37	8 (88.9%)	1 (11.1%)	9	0.0391
	43	11 (92.0%)	1 (8.0%)	12	0.0063
	49	14 (100.0%)	0 (0.0%)	14	0.0001
	55	14 (100.0%)	0 (0.0%)	14	0.0001
	61	17 (100.0%)	0 (0.0%)	17	0.0001
	67	20 (100.0%)	0 (0.0%)	20	<0.0001
	73	19 (95.0%)	1 (5.0%)	20	<0.0001
	79	20 (100.0%)	0 (0.0%)	20	<0.0001
	85	20 (100.0%)	0 (0.0%)	20	<0.0001
Noise	37	4 (50.0%)	4 (50.0%)	8	1.0000
	43	3 (37.5%)	5 (62.5%)	8	0.7266
	49	4 (80.0%)	1 (20.0%)	5	0.3750
	55	5 (62.5%)	3 (37.5%)	8	0.7266
	61	6 (85.7%)	1 (14.3%)	7	0.1250
	67	8 (80.0%)	2 (20.0%)	10	0.1094
	73	19 (100.0%)	0 (0.0%)	19	<0.0001
	79	19 (100.0%)	0 (0.0%)	19	<0.0001
	85	20 (100.0%)	0 (0.0%)	20	<0.0001

3. Phonotaxis scores

An ANOVA revealed significant differences in phonotaxis scores [Fig. 2(d)] across the nine signal levels ($F_{8,288}=27.9$, $P<0.0001$), a significant effect of noise condition ($F_{1,36}=6.2$, $P=0.0173$) and a significant interaction between these two effects ($F_{8,288}=4.8$, $P=0.0002$). No other effects in the model were significant, including the main effect of call order ($F_{1,36}=1.7$, $P=0.2068$) and all interaction terms ($0.3653<Ps<0.6463$). Phonotaxis scores were generally higher for the no-noise group compared with the noise group at signal levels of 67 dB and lower, but only significantly so at the 67-dB signal level [Fig. 2(d)]. The computed sigmoid functions relating mean phonotaxis scores to signal level fit the observed data reasonably well for both the no-noise (adjusted $R^2=0.89$) and noise (adjusted $R^2=0.97$) groups, and most fitted values at a particular signal level fell within one standard error of the observed mean. We summarize in Table II a range of threshold estimates for different threshold criteria expressed as phonotaxis scores (ps') along the fitted sigmoid function. Compared with threshold differences between the no-noise and noise groups based on angular orientation and response probabilities, differences based on phonotaxis scores were much more variable, ranging between 8.9 and 40.4 dB (Table II). In addition, there was no single threshold criterion along the fitted sigmoid curves yielding estimates of threshold for both the no-noise and noise groups (Table II) that were simultaneously consistent with those derived from measures of angular orientation (Sec. III B 1) or from using raw or fitted response probabilities (Sec. III B 2).

IV. EXPERIMENT 2: POPULATION-LEVEL RECOGNITION THRESHOLDS IN NO-CHOICE TESTS

In our second experiment, we used a testing protocol that followed that of Experiment 1 (Sec. III) with one notable exception. In Experiment 2, we used a series of "no-choice" tests in which the same target signal used in Experiment 1 was the only signal presented. Hence, there was no alternating non-target signal, and therefore subjects could not choose between two signals. As in Experiment 1, we regard estimates of recognition thresholds as "population-level" estimates because they are based on the collective responses of a pool of subjects.

A. Methods

1. Experimental design

Aside from the lack of a non-target signal, all experimental details, including the 9 signal level (within subjects) $\times 2$ noise condition (between subjects) factorial design, were the same as those described above for Experiment 1 (Sec. III A 1).

2. Data analyses

Our analyses of the data generally follow those outlined for Experiment 1 (Sec. III A 2). We scored a correct response if subjects touched the arena wall in the 15° arc in front of the speaker broadcasting the target signal. We scored a "no response" if subjects failed to meet this response criterion within 5 min, but responded in the next reference condition. We assessed the directedness of phonotaxis (i.e., angular orientation at a distance of 20 cm from the central release point) using circular statistics (V tests), we compared response probabilities using generalized linear models and nonparametric statistics, and we evaluated differences in phonotaxis scores using ANOVA. We computed estimates of recognition thresholds using the same procedures outlined above for Experiment 1 (Sec. III A 2). Of the 46 females tested in Experiment 2, 6 were excluded from the final data set ($N=40$)

TABLE IV. Results of circular statistical analyses for Experiment 2 (no-choice tests).

Noise condition	Signal condition	Mean vector ($\mu°$)	Length of mean vector (r)	Circular SD (deg)	N	V	P
No-noise	Reference 1	−5	0.86	31	20	0.86	<0.0001
	Reference 2	4	0.91	25	20	0.91	<0.0001
	Reference 3	−2	0.95	18	20	0.95	<0.0001
	Reference 4	0	0.94	20	20	0.94	<0.0001
	37 dB	13	0.27	93	19	0.26	0.0550
	43 dB	44	0.45	72	20	0.32	0.0200
	49 dB	−4	0.72	47	19	0.71	<0.0001
	55 dB	3	0.66	52	20	0.66	<0.0001
	61 dB	0	0.65	53	20	0.65	<0.0001
	67 dB	−6	0.74	44	20	0.74	<0.0001
	73 dB	3	0.90	26	20	0.90	<0.0001
	79 dB	−1	0.93	22	20	0.93	<0.0001
	85 dB	−5	0.89	28	19	0.88	<0.0001
Noise	Reference 1	3	0.91	24	20	0.91	<0.0001
	Reference 2	1	0.97	15	20	0.97	<0.0001
	Reference 3	−1	0.93	23	20	0.93	<0.0001
	Reference 4	−2	0.84	34	20	0.84	<0.0001
	37 dB	−99	0.18	106	20	−0.03	0.5690
	43 dB	−177	0.23	98	20	−0.23	0.9250
	49 dB	−32	0.15	111	20	0.13	0.2060
	55 dB	139	0.03	154	20	−0.02	0.5520
	61 dB	96	0.27	93	20	−0.03	0.5750
	67 dB	154	0.09	126	20	−0.08	0.6930
	73 dB	−19	0.50	68	20	0.47	0.0010
	79 dB	5	0.82	36	20	0.82	<0.0001
	85 dB	−4	0.89	28	20	0.88	<0.0001

because they did not meet our inclusion criteria. None of the subjects in this experiment had been tested previously.

B. Results and discussion

Subjects in the no-noise and noise groups remained similarly motivated to respond to the target signal over the entire duration of the test sequence. Orientation toward the target signal was uniformly strong across all four reference conditions (Table IV). The mean (±SD) latency with which individuals met our response criterion was 71.0 ± 20.6 s, averaged across all four reference conditions and both noise conditions. The mean latencies in the first (65.2 ± 18.7 s), second (73.0 ± 22.3), third (75.7 ± 21.4), and fourth (70.1 ± 19.2) reference conditions differed significantly ($F_{3,114}$=4.2, P=0.0114); however, there was no difference in latency between the no-noise and noise groups ($F_{1,38}$=1.8, P=0.1851), nor was there an interaction between group membership and the sequential order of the reference conditions ($F_{3,114}$=2.1, P=0.1176). In addition, a linear contrast comparing latencies across the sequentially ordered reference conditions was not significant ($F_{1,38}$=3.1, P=0.0880).

1. Angular orientation

An analysis of orientation angles (Table IV) revealed results strikingly similar to those reported above for two-choice tests in Experiment 1 (see Table I). Subjects oriented toward the target signal at signal levels of 43 dB and higher in the no-noise group (Table IV). In the presence of the chorus-shaped noise, signal levels of 73 dB and higher elicited significant orientation toward the target signal. Taking LBs of 37 and 67 dB, we estimated recognition thresholds of 41 and 71 dB in the no-noise and noise groups, respectively, which corresponds to a threshold differences of 30 dB between these two groups.

2. Response probabilities

The proportion of subjects that met our response criteria increased as a function of increasing signal level in both the no-noise group [Fig. 2(e)] and in the noise group [Fig. 2(f)]. Again, this level-dependent increase in responsiveness started at higher signal levels in the noise group compared with the no-noise group [Fig. 2(g)]. The model parameter for signal level was significantly different from zero (χ^2=18.8, P<0.0001, df=1), but those for noise condition (χ^2=0.1, P=0.7431, df=1) and the interaction between signal level and noise condition (χ^2=2.0, P=0.1620, df=1) were not. Subsequent contrast analyses based on estimates of least-squares means revealed that response proportions in the no-noise group differed significantly from zero (χ^2=31.1, P<0.0001, df=1), while those in the noise group did not (χ^2=0.41, P=0.5210, df=1). In addition, there were significant differences between the proportions responding in the no-noise and noise groups (χ^2=22.9, P<0.0001, df=1). Table V summarizes threshold estimates as a function of different threshold criteria expressed as response probabilities (p') along best-fit logistic regression curves [Fig. 2(g)]. As in

TABLE V. Behavioral recognition thresholds in the no-noise and noise groups of Experiment 2 (no-choice tests) as functions of threshold criteria based on fitted response probabilities and phonotaxis scores.

Response variable	Threshold criterion	Estimated threshold signal levels (dB SPL)		Threshold difference (dB)
		No-noise	Noise	
Response probability (p')	0.2	23.6	41.2	17.6
	0.3	28.1	47.9	19.8
	0.4	31.8	53.4	21.6
	0.5	35.1	58.4	23.3
	0.6	38.5	63.4	24.9
	0.7	42.2	68.9	26.7
	0.8	46.7	75.6	28.9
	0.9	53.4	85.6	32.2
Phonotaxis scores (ps')	0.2	33.5	61.9	28.4
	0.3	35.9	64.7	28.8
	0.4	37.9	66.2	28.3
	0.5	39.7	67.4	27.7
	0.6	41.5	68.5	27.0
	0.7	43.3	69.6	26.3
	0.8	45.2	70.7	25.5
	0.9	47.4	72.1	24.7

Experiment 1, there was no single criterion that produced threshold estimates for both the no-noise and noise groups that were consistent with those derived above based on angular orientations (Sec. IV B 1) and raw response probability (see below, this section). In contrast to Experiment 1, however, there was considerably more variation in estimates of threshold differences, which ranged between 17.7 and 32.3 dB (Table V; cf Table II).

The lowest signals level at which at least 50% of subjects responded was 43 dB (15 of 20 responded) in the no-noise group and 67 dB (12 of 20 responded) in the noise group, with corresponding LBs of 37 and 61 dB. These UBs and LBs yielded threshold estimates of 41 and 65 dB in the no-noise and noise groups, respectively, representing a threshold difference of 24 dB between the two groups. This magnitude of difference is smaller than the 30-dB difference determined in parallel analyses of results from the two-choice tests of Experiment 1 (Sec. III B 2). In the no-choice tests of Experiment 2, the proportion of subjects in the no-noise group that responded was significantly higher than the expected false alarm rate of 0.20 at the 37-dB signal level (at which 10 of 20 subjects responded) and all higher levels [Fig. 2(e)]. We again assumed 37 dB to be an UB and calculated a threshold of 35 dB for this group. Parallel analyses for the noise group yielded an UB and a LB of 67 and 61 dB, respectively, and a threshold estimate of 65 dB. Thus, estimates based on statistically significant differences from the expected false alarm rate yielded the same absolute thresholds, and thus the same threshold difference (30 dB), as in the two-choice tests described earlier (Sec. III B 2).

3. Phonotaxis scores

As in Experiment 1, phonotaxis scores increased as a function of increasing signal level, and this level-dependent increase began at higher signal levels in the noise group compared with the no-noise group [Fig. 2(h)]. There were significant main effects of signal level ($F_{8,304}=26.4$, $P<0.0001$) and noise condition ($F_{1,38}=45.1$, $P<0.0001$) and a significant interaction between these two effects ($F_{8,304}=7.8$, $P<0.0001$). Phonotaxis scores were similar between the two noise conditions at signal levels of 37, 79, and 85 dB, but they were significantly higher in the no-noise group at all other signal levels [Fig. 2(h)]. The fitted sigmoid relationships between mean phonotaxis scores and signal level explained large portions of the variance in both the no-noise (adjusted $R^2=0.86$) and noise (adjusted $R^2=0.96$) groups. Most fitted values at each nominal signal level fell within one standard error of the actual mean phonotaxis score observed at that level. Table V summarizes threshold estimates as a function of different threshold criteria expressed as phonotaxis scores (ps') along the fitted sigmoid functions. There was generally less variation in the threshold differences between the no-noise and noise groups (24–29 dB, Table V) compared with those from parallel analyses of two-choice tests in Experiment 1 (8–41 dB, Table II). Again, however, no single threshold criterion based on fitted phonotaxis scores yielded threshold estimates for both the no-noise and the noise groups that were entirely consistent with those derived above for angular orientation (Sec. IV B 1) and response probabilities (Sec. IV B 2).

V. EXPERIMENT 3: INDIVIDUAL-LEVEL RECOGNITION THRESHOLDS IN NO-CHOICE TESTS

In Experiment 3, we used no-choice tests and an adaptive tracking method of threshold estimation based loosely on the method of limits commonly used to determine thresholds in more traditional psychoacoustic studies (Klump *et al.*, 1995). Each subject was again tested at several signal levels, but the levels chosen for all but the first test were contingent upon the subject's response in the previous test. In this way, we were able to derive threshold estimates for each individual separately.

A. Methods

The target signal and chorus-shaped noises were the same as those described above for Experiment 1 (Sec. III A 1) and Experiment 2 (Sec. IV A 1). We randomly assigned subjects either to a no-noise group ($N=20$) or to a noise group ($N=20$) for which chorus-shaped noise was broadcast continuously during a test from an overhead speaker at a long-term overall SPL of 70 dB (LC_{eq}). None of the subjects in this experiment had been tested previously.

For each individual subject, a test sequence comprised a variable number of reference conditions and treatment conditions that was determined by the subject's responses. Again, we scored a correct response if subjects touched the wall of the test arena in the 15° arc in front of the target speaker in under 5 min. We recorded a no response when subjects failed to meet this criterion. Each sequence began and ended with the reference condition, which again consisted of the target signal presented alone at 85 dB SPL. We also tested the reference condition after any two consecutive

treatment conditions failed to elicit correct responses. The first treatment condition in the sequence always involved presenting the target signal at a level estimated by us to be close to the recognition threshold in the noise condition being tested. Our initial estimates were based on results reported by Beckers and Schul (2004) and Bee and Swanson (2007) and were determined *prior to* any analyses of results from Experiments 1 and 2. For the no-noise group, the initial signal level was 45 dB for the first three subjects tested but was subsequently reduced to 39 dB for the remaining 17 subjects. For the group tested with chorus-shaped noise, the initial signal level was 70 dB for all subjects. Following the first treatment condition, we reduced or increased the signal level by 3 dB in the subsequent treatment condition depending on whether the subject did or did not respond in the previous treatment condition, respectively. We continued either decreasing or increasing the signal level in 3-dB steps in subsequent treatment conditions until the subject's behavior changed (e.g., going from correct response to incorrect response between two consecutive treatment conditions). After the subject's behavior changed, we tested a final treatment condition in which we reversed the direction of signal level change by a reduced step-size of 1.5 dB. If the subject responded in this final treatment condition, the signal level for that treatment condition was used as the UB of a threshold estimate, and the next lowest level tested was used as the LB. If the subject failed to respond in the final treatment condition, the signal level in that condition was taken as the LB of the threshold and the next highest signal level previously eliciting a response was taken as the UB. We computed an estimate of the recognition threshold as the average of the UB and LB using Eq. (1) and compared these between groups using a Mann-Whitney U Test.

B. Results and discussion

The mean (±SD) latencies to respond to the target signal in the first (80.7 ± 24.4 s; $N=40$) and last (86.9 ± 29.4 s; $N=40$) reference conditions did not differ significantly ($F_{1,38}=2.3$, $P=0.1359$). There were no differences in response latency between subjects in the no-noise and noise groups ($F_{1,38}=0.4$, $P=0.5532$), nor was there an interaction between noise condition and reference condition ($F_{1,38}=1.1$, $P=0.2965$). The numbers of signal levels at which individuals in the two noise conditions were tested ranged between 3 and 6 levels and did not differ between groups (Mann-Whitney U Test: $U=190$, $P=0.7868$); the median and the modal number of signal levels tested were 3 levels in both noise groups.

The difference between the median thresholds determined for the no-noise and noise groups was 32.5 dB (Fig. 3) and was statistically significant (Mann-Whitney U Test: $U=0.00$, $P<0.0001$). Across subjects tested in the no-noise group, the LBs of threshold estimates ranged between 31.5 and 42.0 dB and the UBs ranged between 33.0 and 43.5 dB. The median threshold for subjects in the no-noise group was 38.3 dB (Fig. 3). In the noise group, the median threshold was 70.8 dB. Across individuals assigned to the noise group, LBs ranged between 62.5 and 76 dB and UBs ranged between 64.0 and 77.5 dB.

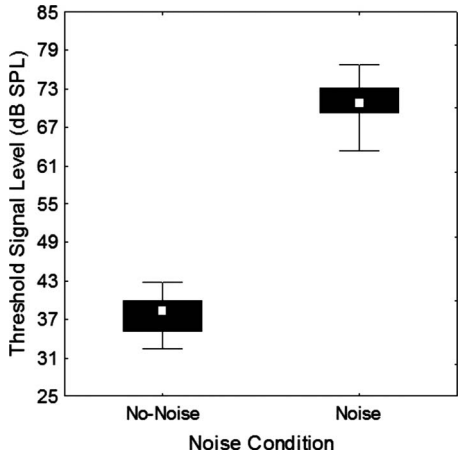


FIG. 3. Individual recognition thresholds based on no-choice tests in the no-noise and noise groups of Experiment 3. Depicted here are the median (point), inter-quartile range (box), and range (whiskers).

VI. GENERAL DISCUSSION

One goal of previous studies of gray treefrogs (Gerhardt, 2001) and indeed many other anurans (Gerhardt and Huber, 2002) has been to quantify the signal properties that elicit recognition of sound patterns as the sexual advertisement signal of an appropriate mate. Two related issues in the literature concern how sound pattern recognition is defined and experimentally measured (e.g., Ryan and Rand, 1993, 2001; Bush *et al.*, 2002) and how the mechanisms for sound pattern recognition operate in the face of constraints posed by noisy environments (Gerhardt and Klump, 1988; Feng and Ratnam, 2000; Feng and Schul, 2007). Comparatively few studies have explicitly investigated sound pattern recognition as a function of signal level under conditions that also included controlled exposure to natural or artificial sources of noise designed to simulate the acoustic environment of a breeding chorus (but see Ehret and Gerhardt, 1980; Gerhardt and Klump, 1988; Schwartz and Gerhardt, 1998; Wollerman, 1999; Schwartz *et al.*, 2001; Wollerman and Wiley, 2002; Bee, 2007, 2008a, 2008b). Our aim in this study was to evaluate several empirical and analytical methods for estimating sound pattern recognition thresholds in frogs using phonotaxis as a behavioral assay.

A. Phonotaxis as a behavioral response measure

In contrast to studies of most other vertebrates (e.g., Fay and Popper, 1999; Dooling *et al.*, 2000), traditional psychoacoustic approaches based on classical or operant conditioning are notoriously difficult or unsuccessful in studies of anuran hearing (but see Elepfandt *et al.*, 2000). While related experimental procedures, such as reflex modification (reviewed in Simmons and Moss, 1995), have met with some success, they have not been widely adopted. Phonotaxis assays remain the most common experimental approach used to address questions about frog hearing and acoustic communication (reviewed in Gerhardt and Huber, 2002).

Among the advantages of phonotaxis as a behavioral assay are that it can be used to exploit the animal's natural behavioral repertoire to address ecologically valid questions. What is more, many frog species reliably exhibit robust pho-

TABLE VI. Comparison of various methods for estimating call recognition thresholds from this experiment and two other studies.

Experiment or published study	Method of threshold estimation	Threshold (dB SPL)		Threshold difference (dB)
		No-noise	Noise	
Experiment 1 – Two-choice tests	Angular orientation at 20 cm: average of lowest signal level yielding significant orientation and next lowest level	41	71	30
	Response probability ($p>0.2$): average of lowest signal level yielding significant result in a binomial test of $p>0.2$ and next lowest level	35	65	30
	Response probability ($p>0.5$): average of lowest signal level yielding $p>0.5$ and next lowest level	41	71	30
	Choice results: average of lowest signal level yielding significant preference ($p>0.5$) for conspecific calls over heterospecific calls	35	71	36
	Response probability ($p'=0.50$): interpolated value using threshold criterion of $p'=0.50$ from logistic regression equations	42	62	20
	Phonotaxis scores ($ps'=0.50$): interpolated value using threshold criterion of 0.50 from fitted sigmoid functions	47	68	21
Experiment 2 – No-choice tests	Angular orientation at 20 cm: average of lowest signal level yielding significant orientation and next lowest level	41	71	30
	Response probability ($p>0.2$): average of lowest signal level yielding significant result in a binomial test of $p>0.2$ and next lowest level	35	65	30
	Response probability ($p>0.5$): average of lowest signal level yielding $p>0.5$ and next lowest level	41	65	24
	Response probability ($p'=0.50$): interpolated value using threshold criterion of $p'=0.50$ from logistic regression equations	35	58	23
	Phonotaxis scores ($ps'=0.50$): interpolated value using threshold criterion of 0.50 from fitted sigmoid functions	40	67	27
Experiment 3 – No-choice tests	Adaptive tracking	38	71	33
Bee and Swanson (2007) – No-choice tests	Response probability ($p'=0.50$): interpolated value using threshold criterion of $p'=0.50$ from logistic regression equation	42	68	26
Beckers and Schul (2004) – No-choice tests	Response probability ($p>0.5$): average of lowest signal level yielding $p>0.5$ and next lowest level	41	⋯	⋯

notaxis under highly controlled laboratory conditions. Among the disadvantages of phonotaxis as a behavioral assay is that it does not (and cannot) distinguish between signal detection and signal recognition. This follows because a female might fail to exhibit phonotaxis either because she could not detect the sound or because she detected the sound but did not recognize it as the signal of an appropriate mate. Using phonotaxis as a behavioral measure of sound pattern recognition is further complicated by the fact that recognition in some species is not an "all-or-none" phenomenon, but instead may be a continuous function of variation in one or more signal attributes unrelated to signal amplitude (Bush *et al.*, 2002). In addition, phonotaxis behavior cannot be used to generate robust data on difference limens because *perceptual* discrimination between two stimuli may be possible even though individuals exhibit no *behavioral* discrimination. In summary, phonotaxis is a useful tool for investigating "just meaningful differences" (Nelson and Marler, 1990) but cannot by itself provide information on "just noticeable differences." With these caveats in mind, we evaluated several experimental and analytical approaches for deriving estimates of signal recognition thresholds in the presence and absence of masking noise using phonotaxis as a behavioral assay.

B. Comparing behavioral measures of signal recognition thresholds

In studies of humans, the SRT depends on the rate of correct responses (i.e., correctly recognizing a spoken word in masking noise and not simply recognizing that a word was spoken). Therefore, we similarly limited our analyses to correct responses by operationally defining signal recognition as occurring when females exhibited phonotaxis in response to a conspecific advertisement call. Our most striking finding is the generally high degree of agreement among estimates of both absolute thresholds and relative differences derived using different experimental methods and analytical approaches (Table VI). For example, recognition thresholds derived from measures of angular orientation were exactly the same in Experiment 1 (two-choice tests) and Experiment 2 (no-choice tests). Likewise, threshold estimates based on significant differences from an expected false alarm rate ($p=0.20$) were identical between Experiments 1 and 2 and were within one step-size in signal level (6 dB) of those

based on angular orientation for both noise conditions. Threshold estimates based on the signal levels at which the raw proportion of females exhibiting correct responses exceeded 50% were identical for the no-noise groups of Experiments 1 and 2, and they were within one signal-level step-size (6 dB) of each other in the noise groups of these two experiments. Threshold estimates based on the probability of choosing a conspecific over a heterospecific call were also similar to these other estimates. Perhaps most importantly, all estimates of absolute thresholds in Experiments 1 and 2 that were based on angular orientation or raw response probabilities were within 0–6 dB of those derived in Experiment 3 using smaller step-sizes (1.5 and 3 dB) and an entirely different approach based on adaptive tracking. Moreover, estimates of absolute recognition thresholds are in generally good agreement with those reported in previously published studies of gray treefrogs (Table VI). With only one exception (Table VI), the magnitudes of threshold difference between the two noise conditions based on angular orientation and raw probabilities in Experiments 1 and 2 were within 3 dB of that derived in Experiment 3. In general, threshold estimates based on fitted response probabilities (logistic functions) and phonotaxis scores (sigmoid functions) were similar between Experiments 1 and 2, but they were also much more variable and tended to yield smaller threshold differences (≈20–27 dB; Table VI) between the two noise conditions compared to our other estimates.

Given the general similarity among the results from Experiments 1–3, it is worth considering practical and logistical differences between them. The adaptive tracking procedure we used in Experiment 3 had a number of advantages over the methods we used for Experiments 1 and 2. First, by allowing us to estimate a threshold for each individual, our approach in Experiment 3 allowed us to generate measures of central tendency and variability for each noise condition. These measures, in turn, allowed us to make a straightforward between-groups statistical comparison of recognition thresholds in the two noise conditions. Such a direct comparison was not possible in Experiments 1 and 2. Second, our estimates of threshold in Experiment 3 were less dependent on sample size. Some of the upper and lower threshold bounds in Experiments 1 and 2 depended on cutoffs based on level-dependent patterns of statistical significance, which would vary as a function of statistical power, and hence sample size, for any given effect size. Third, we were able to test smaller gradations in signal level in Experiment 3 by using smaller step-sizes (e.g., 1.5–3 dB) compared with those in Experiments 1 and 2 (6 dB), which, in turn, might provide for better accuracy and precision in estimates of recognition thresholds. Fourth, our approach in Experiment 3 required less time and fewer tests of each subject (e.g., three signal levels) compared with our approach in Experiments 1 and 2, in which each subject was tested at all possible signal levels (e.g., nine signal levels). Finally, our approach in Experiment 3 required relatively fewer subjects; no females failed to complete Experiment 3, whereas 13% (12 of 92) of the females tested in Experiments 1 and 2 failed to complete the whole series of tests. The relative advantages of Experiment 3 over Experiments 1 and 2 in terms of shorter testing times and smaller subject pools might be diminished if the approach used in Experiments 1 and 2 were modified so that a test ended as soon as the subject advanced 20 cm from the release point. Our results suggest that such a modified method would yield results similar to the adaptive tracking procedure of Experiment 3.

VII. CONCLUSION

Efforts to understand the mechanisms by which humans understand speech in noisy social settings hold a central place in modern hearing research. One prominent research methodology involves estimating SRTs in various masking conditions to understand how the spectral, temporal, and spatial relationships between sources of signals and sources of masking noise influence speech perception. Similar experimental approaches have not been widely adopted in studies aimed at discovering how various nonhuman animals have evolved to solve similar "cocktail-party-like" communication problems. Given the exceptional value of anuran amphibians as model systems for studying the mechanisms of acoustic communication in noisy environments, this study aimed to compare estimates of signal recognition thresholds using several common experimental methods and analytical tools. Our results reveal insights into how phonotaxis experiments might best be used to answer questions concerning how frogs recognize behaviorally relevant sound patterns in high levels of biologically realistic background noise. The main conclusion of this study is that the methods and analyses compared here yielded generally quite similar results; however, they differed in a number of practical ways that will be important to consider in designing future experiments.

ACKNOWLEDGMENTS

We extend special thanks to D. Heil, J. Henderson, J. Henly, M. Kuczynski, J. Lane, A. Leightner, E. Love, K. Riemersma, D. Rittenhouse, K. Speirs, S. Tekmen, A. Thompson, A. Vélez, and J. Walker-Jansen for their assistance in collecting and testing frogs and especially to S. Tekmen for her extensive help compiling data and A. Vélez for assistance with stimulus generation. This work was conducted under Special Use Permit No. 14902 from the Minnesota Department of Natural Resources and with a Special Use Permit issued by M. Linck at the Three Rivers Park District. This research was approved by the University of Minnesota Institutional Animal Care and Use Committee (Protocol No. 0809A46721) and was funded by a Grant-in-Aid of Research from the University of Minnesota Graduate School and NIH Grant No. R03-DC008396 to M.B. and NSF Grant No. IBN-0342183 to J.S.

Aubin, T., and Jouventin, P. (**2002**). "How to vocally identify kin in a crowd: The penguin model," Adv. Study Behav. **31**, 243–277.

Beckers, O. M., and Schul, J. (**2004**). "Phonotaxis in *Hyla versicolor* (Anura, Hylidae): The effect of absolute call amplitude," J. Comp. Physiol. [A] **190**, 869–876.

Bee, M. A. (**2007**). "Sound source segregation in grey treefrogs: Spatial release from masking by the sound of a chorus," Anim. Behav. **74**, 549–558.

Bee, M. A. (**2008a**). "Finding a mate at a cocktail party: Spatial release from masking improves acoustic mate recognition in grey treefrogs," Anim.

Behav. **75**, 1781–1791.
Bee, M. A. (**2008b**). "Parallel female preferences for call duration in a diploid ancestor of an allotetraploid treefrog," Anim. Behav. **76**, 845–853.
Bee, M. A., and Micheyl, C. (**2008**). "The cocktail party problem: What is it? How can it be solved? And why should animal behaviorists study it?," J. Comp. Psychol. **122**, 235–251.
Bee, M. A., and Swanson, E. M. (**2007**). "Auditory masking of anuran advertisement calls by road traffic noise," Anim. Behav. **74**, 1765–1776.
Bush, S. L., Gerhardt, H. C., and Schul, J. (**2002**). "Pattern recognition and call preferences in treefrogs (Anura: Hylidae): A quantitative analysis using a no-choice paradigm," Anim. Behav. **63**, 7–14.
Cherry, E. C. (**1953**). "Some experiments on the recognition of speech, with one and with two ears," J. Acoust. Soc. Am. **25**, 975–979.
Dooling, R. J., Fay, R. R., and Popper, A. N. (**2000**). *Comparative Hearing: Birds and Reptiles* (Springer, New York).
Ehret, G., and Gerhardt, H. C. (**1980**). "Auditory masking and effects of noise on responses of the green treefrog (*Hyla cinerea*) to synthetic mating calls," J. Comp. Physiol. [A] **141**, 13–18.
Elepfandt, A., Eistetter, I., Fleig, A., Gunther, E., Hainich, M., Hepperle, S., and Traub, B. (**2000**). "Hearing threshold and frequency discrimination in the purely aquatic frog *Xenopus laevis* (pipidae): Measurement by means of conditioning," J. Exp. Biol. **203**, 3621–3629.
Fay, R. R., and Popper, A. N. (**1999**). *Comparative Hearing: Fish and Amphibians* (Springer, New York).
Feng, A. S., and Ratnam, R. (**2000**). "Neural basis of hearing in real-world situations," Annu. Rev. Psychol. **51**, 699–725.
Feng, A. S., and Schul, J. (**2007**). "Sound processing in real-world environments," in *Hearing and Sound Communication in Amphibians*, edited by P. A. Narins, A. S. Feng, R. R. Fay, and A. N. Popper (Springer, New York), pp. 323–350.
Gerhardt, H. C. (**1975**). "Sound pressure levels and radiation patterns of vocalizations of some North American frogs and toads," J. Comp. Physiol. **102**, 1–12.
Gerhardt, H. C. (**1995**). "Phonotaxis in female frogs and toads: Execution and design of experiments," in *Methods in Comparative Psychoacoustics*, edited by G. M. Klump, R. J. Dooling, R. R. Fay, and W. C. Stebbins (Birkhäuser, Basel), pp. 209–220.
Gerhardt, H. C. (**2001**). "Acoustic communication in two groups of closely related treefrogs," Adv. Stud. Behav. **30**, 99–167.
Gerhardt, H. C., and Huber, F. (**2002**). *Acoustic Communication in Insects and Anurans: Common Problems and Diverse Solutions* (Chicago University Press, Chicago).
Gerhardt, H. C., and Klump, G. M. (**1988**). "Masking of acoustic signals by the chorus background noise in the green treefrog: A limitation on mate choice," Anim. Behav. **36**, 1247–1249.
Gerhardt, H. C., Ptacek, M. B., Barnett, L., and Torke, K. G. (**1994**). "Hybridization in the diploid-tetraploid treefrogs *Hyla chrysoscelis* and *Hyla versicolor*," Copeia **1994**, 51–59.
Greenhouse, S. W., and Geisser, S. (**1959**). "On methods in the analysis of profile data," Psychometrika **24**, 95–112.
Holloway, A. K., Cannatella, D. C., Gerhardt, H. C., and Hillis, D. M. (**2006**). "Polyploids with different origins and ancestors form a single sexual polyploid species," Am. Nat. **167**, E88–E101.
Horton, N. J., and Lipsitz, S. R. (**1999**). "Review of software to fit generalized estimating equation regression models," Am. Stat. **53**, 160–169.
Klump, G. M., Dooling, R. J., Fay, R. R., and Stebbins, W. C. (**1995**). *Methods in Comparative Psychoacoustics* (Birkhäuser, Basel).
Narins, P. M., and Zelick, R. (**1988**). "The effects of noise on auditory processing and behavior in amphibians," in *The Evolution of the Amphibian Auditory System*, edited by B. Fritzsch, M. J. Ryan, W. Wilczynski, T. E. Hetherington, and W. Walkowiak (Wiley, New York), pp. 511–536.
Nelson, D. A., and Marler, P. (**1990**). "The perception of birdsong and an ecological concept of signal space," in *Comparative Perception*, edited by M. A. Berkley and W. C. Stebbins (Wiley, New York), Vol. **II**, pp. 443–478.
Plomp, R. (**1978**). "Auditory handicap of hearing impairment and limited benefit of hearing aids," J. Acoust. Soc. Am. **63**, 533–549.
Plomp, R., and Mimpen, A. M. (**1979a**). "Improving the reliability of testing the speech reception threshold for sentences," Audiology **18**, 43–52.
Plomp, R., and Mimpen, A. M. (**1979b**). "Speech reception threshold for sentences as a function of age and noise level," J. Acoust. Soc. Am. **66**, 1333–1342.
Ryan, M. J., and Rand, A. S. (**1993**). "Species recognition and sexual selection as a unitary problem in animal communication," Evolution **47**, 647–657.
Ryan, M. J., and Rand, A. S. (**2001**). "Feature weighting in signal recognition and discrimination by túngara frogs," in *Anuran Communication*, edited by M. J. Ryan (Smithsonian Institution, Washington, DC), pp. 86–101.
Schul, J., and Bush, S. L. (**2002**). "Non-parallel coevolution of sender and receiver in the acoustic communication system of treefrogs," Proc. R. Soc., London, Ser. B **269**, 1847–1852.
Schwartz, J. J., Buchanan, B. W., and Gerhardt, H. C. (**2001**). "Female mate choice in the gray treefrog (*Hyla versicolor*) in three experimental environments," Behav. Ecol. Sociobiol. **49**, 443–455.
Schwartz, J. J., and Freeberg, T. M. (**2008**). "Acoustic interaction in animal groups: Signaling in noisy and social contexts—Introduction," J. Comp. Psychol. **122**, 231–234.
Schwartz, J. J., and Gerhardt, H. C. (**1998**). "The neuroethology of frequency preferences in the spring peeper," Anim. Behav. **56**, 55–69.
Simmons, A. M. and Moss, C. F. (**1995**). "Reflex modification: A tool for assessing basic auditory function in anuran amphibians," in *Methods in Comparative Psychoacoustics*, edited by G. M. Klump, R. J. Dooling, R. R. Fay, and W. C. Stebbins (Birkhäuser, Basel), pp. 197–208.
Swanson, E. M., Tekmen, S. M., and Bee, M. A. (**2007**). "Do female anurans exploit inadvertent social information to locate breeding aggregations?" Can. J. Zool. **85**, 921–932.
Wollerman, L. (**1999**). "Acoustic interference limits call detection in a Neotropical frog *Hyla ebraccata*," Anim. Behav. **57**, 529–536.
Wollerman, L., and Wiley, R. H. (**2002**). "Background noise from a natural chorus alters female discrimination of male calls in a Neotropical frog," Anim. Behav. **63**, 15–22.
Zar, J. H. (**1999**). *Biostatistical Analysis* (Prentice-Hall, Upper Saddle River, NJ).

Linear behavior of a preformed microbubble containing light absorbing nanoparticles: Insight from a mathematical model

E. Sassaroli, K. C. P. Li, and B. E. O'Neill[a)]
Department of Radiology, the Methodist Hospital Research Institute, Houston, Texas 77030

(Received 14 May 2009; revised 21 August 2009; accepted 4 September 2009)

Microbubbles are used as ultrasonic contrast agents in medical imaging because of their highly efficient scattering properties. Gold nanoparticles absorb specific wavelengths of optical radiation very effectively with the subsequent generation of thermo-acoustic waves in the surrounding medium. A theoretical and numerical analysis of the possibility of inducing radial oscillations in a pre-existing spherical microbubble, through the laser excitation of gold nanoparticles contained within, is presented. A description of such a system can be obtained in terms of a confined two-phase model, with the nanoparticles suspended in a confined region of gas, surrounded by a liquid. The Rayleigh–Plesset equation is assumed to be valid at the boundary between the gas and the liquid. The confined two-phase model is solved in linear approximation. The system is diagonalized and the general solution is obtained. This solution is in the form of exponentially decaying oscillatory functions for the temperature and pressure inside the bubble, and radial oscillations of the bubble boundary. It was found that, for the right size of bubbles, the oscillatory behavior takes place in the low megahertz range, which is ideal for medical applications. This study suggests the possibility of new applications of microbubbles in photoacoustic imaging.
 [DOI: 10.1121/1.3238262]

PACS number(s): 43.80.Qf, 43.35.Ud, 43.35.Ei [CCC] Pages: 2802–2813

I. INTRODUCTION

It is well known that small gas bubbles in liquids can be easily made to oscillate upon exposure to sound waves. As a result, gas bubbles are excellent scatterers of sound waves and this property finds applications in very diverse fields such as oceanography, industry, and medicine.[1] In medicine, ultrasound contrast agents, which are gas microbubbles encapsulated by a thin shell, are employed to increase the scattering of ultrasound by blood and thus permit a more efficient visualization of vessels and flowing blood.[2] These microbubbles consist of a perfluorocarbon gas and are coated by a thin shell to avoid clustering and to prevent surface tension from dissolving them. For acoustic waves of small amplitude, the echoes are produced by bubbles resonant at the driving frequency (1–10 MHz in conventional medical ultrasound).

An alternative way to achieve contrast in an ultrasonic image is to use laser excitation of thermo-acoustic pulses, which are then detected with an ultrasound receiver. In medical imaging, this is most often called photoacoustic imaging. Current promising applications of photoacoustic imaging include small animal imaging, tomographic imaging of the skin, and other superficial organs and breast cancer detection by near-infrared light.[3] Image contrast is caused by the local optical absorption properties of the tissue and depends on the frequency content of the resulting wave. This latter arises because of the bandwidth limitations of the receiving transducer as well as the frequency dependent attenuation of tissue. A drawback of this modality is that tissue strongly absorbs/scatters visible light and even infrared light can only penetrate a few centimeters in the tissue. For many years, researchers have worked to enhance the photoacoustic signal through the use of light absorbing dyes and particles (photoacoustic contrast agents), including the use of gold nanoparticles. Gold nanoparticles show an intense absorption of specific frequencies of optical radiation owing to their surface plasmon resonance (SPR) frequency.[4,5] SPR is caused by the collective excitations of the conduction electrons under light irradiation. Because of their intense absorption of optical radiation, gold nanoparticles are under investigation for a variety of biomedical applications,[6–8] such as photothermal therapy for cancer treatment and as contrast enhancement agent in optical imaging modalities. Unfortunately, the use of these agents does not appear to give sufficient enhanced contrast for practical imaging of deep structures. This result might be dramatically improved if the frequency content of the resulting ultrasonic wave could be better matched to the receiving transducer. As it currently stands, this frequency content is determined primarily by the local thermal properties of the tissue and the laser pulse parameters, the latter being limited by safety considerations. Our goal is to explore the practical possibility of designing a much stronger photoacoustic contrast agent by combining light absorbing nanoparticles with microbubbles that resonate primarily at the frequency range of interest. To the best of our knowledge, this is the first study that examines the possibility of placing the gold nanoparticle heat sources inside the gas structure. Several papers can be found in the

[a)]Author to whom correspondence should be addressed. Electronic mail: beoneill@tmhs.org

literature that investigate the process of microbubble nucleation and dynamics around gold nanoparticles.[9–11] Such microbubbles are explored for cancer cell killing,[12,13] the externally triggered release of drugs from a liposome,[14] and as imaging contrast agents.[15]

For *in vivo* applications, the potential advantages of having gold nanoparticles enclosed in a preformed microbubble rather than just simply generating them by laser irradiation are increased photoacoustic efficiency, decreased energy dose of laser light, and reduction in the potential for damage to tissues. A major and challenging requirement for the successful application of such an idea is the ability of engineering a well-designed microbubble that embeds gold nanoparticles in it through the use of ligands. One important difference between the two approaches is that single gold nanoparticles can extravasate, for example, into the interstitial space of tumors that have leaky vessels. On the other hand, if gold nanoparticles are embedded in the microbubble, they will remain in the blood vessels. Such microbubbles are easily combined with ligands that could potentially be used to target them to various pathologies, for example, the angiogenic regions of cancers, for diagnostic or treatment purposes. Current ultrasound technology does not have the sensitivity to detect these few microbubbles with any reliability against a tissue background. Using light to excite acoustic emissions in these microbubbles by way of embedded light absorbing gold nanoparticles is hopefully a way to enhance this sensitivity. The result would be a reliable but inexpensive molecular imaging technology. Since differently sized nanoparticles respond to different wavelengths of light, an added bonus should be the ability to detect various molecular targets in one single doctor visit, contrary to positron emission tomography that requires more than one appointment.

This work represents a first step in testing the hypothesis that embedding the nanoparticles inside the liposome or polymer structure containing gas will result in advantages sufficient to justify the engineering effort required.

A mathematical description of such a system can be developed in terms of a confined two-phase model with the solid particles (gold nanoparticles) in suspension in a fluid phase, i.e., the gas. The gas is assumed to be confined in a bubble surrounded by water. The effect of the shell that is normally present in ultrasound contrast agent is not considered in this study. The suspension is assumed to be diluted so that nanoparticle-nanoparticle interaction can be ignored. The problem can be solved analytically only in linear approximation. Even in linear approximation, however, the task is rather complex and several approximations have been made in this study. Because we are interested to the heat exchange between the particles and the gas, it is assumed that the gas and the particles have the same velocity and pressure. The pressure and the temperatures of the two phases are assumed to be functions of time only. The gas is assumed ideal and the gas/vapor diffusion equation is disregarded. The heat exchange between the gas and the nanoparticles is treated using a low frequency heat-transfer expression. Possible quantum mechanical corrections to this expression are not considered in this paper. Gravity effects are also ignored. Furthermore, it is assumed that the bubble is in an unbounded liquid and the linearized Rayleigh–Plesset equation is used to describe the bubble's radial oscillations in water. In an *in vivo* situation, microbubbles are confined in a blood vessel and therefore the above unbounded boundary condition in the liquid does not hold.[16] Although in principle the introduction of boundaries represented by the vessel can be implemented in a straightforward way, the problem of the bubble dynamics will become much more complex and its solution would require computer simulations. In addition, the linear approximation for the bubble dynamics can hold only for sufficiently weak laser pulses. As the laser power is increased, the oscillations will become non-linear. Once again, the solution of the system would require computer simulations. Several computational techniques that deals with this problem are available, for example, the free-Lagrange numerical method[17] or the coupled-finite element and boundary element method.[18]

Within the limitations described above, the model can predict the radial oscillations of the bubble containing the gold nanoparticle suspension, the gas pressure, and the temperature of the two phases as a function of time and exposure conditions. It also shows that its oscillation frequency is in the megahertz range and it is therefore an efficient generator of ultrasound for biomedical applications. From a mathematical point of view, this model provides a framework upon which more complete photoacoustic models of a preformed gas bubble containing light absorbing nanoparticles could be developed.

For the purpose of comparison, we have also considered the situation in which the fluid inside the micro-structure containing gold nanoparticles is a liquid and not an ideal gas. We have considered a linearized equation of state for the liquid, which for simplicity we assume to be water. In such a situation, it can be seen that the radial oscillations are very weak and that the natural frequency of oscillation for such a system is impractically high for medical purposes.

II. MATHEMATICAL FORMULATION

A. General equations

Two-phase models have been investigated by several authors (see, for example, Refs. 19–26). In this study, we have adopted the formulation developed by Temkin,[20] which is particularly suited for suspensions. In two-phase models, it is assumed that each phase obeys its own set of conservation equations. Because we are primarily interested in the heat exchange between the particles and the gas, we have made the simplification that the gas and the particles have the same velocity and pressure. In such a situation, one needs to consider only the equations of continuity and energy for both the particulate phase and the gas phase. The momentum equations can be substituted with the relation

$$p = p(t), \tag{1}$$

which describes the common pressure for the suspension. The continuity equation for the fluid phase and the particular phase are, respectively,

$$\frac{D\sigma_f}{Dt} + \sigma_f \nabla \cdot \mathbf{u} = 0, \tag{2}$$

$$\frac{D\sigma_p}{Dt} + \sigma_p \nabla \cdot \boldsymbol{u} = 0, \tag{3}$$

where $\mathbf{u}$ is the common velocity field, σ_f is the density of the gas phase, and σ_p is the density of the particular phase, defined as

$$\sigma_f = \rho_f(1 - \phi_v(\mathbf{r},t)), \quad \sigma_p = \rho_p \phi_v(\mathbf{r},t), \tag{4}$$

where ρ_f and ρ_p are the densities of the gas and the particle material, with ρ_p constant for rigid particles. ϕ_v denotes the particle volume concentration with $\phi_v = n v_p$, where n indicates the number of particles per unit volume and v_p is the particle volume. The symbol D/Dt represents the convective derivative $D/Dt = (\partial/\partial t) + \mathbf{u} \cdot \nabla$. The energy equation for the gas phase is

$$\sigma_f \left(c_{pf} \frac{DT_f}{Dt} - \frac{\beta_f T_f}{\rho_f} \dot{p} \right) = \dot{q}_p, \tag{5}$$

with c_{pf} is the gas specific heat at constant pressure, β_f is the gas thermal expansion coefficient, T_f is the gas phase temperature, and $\dot{q}_p$ is the distributed particle heat-transfer rate per unit volume of suspension. The energy equation for the particulate phase assuming the particles to be rigid is

$$\sigma_p c_{pp} \frac{DT_p}{Dt} = -\dot{q}_p + \dot{q}_{\text{laser}}, \tag{6}$$

where c_{pp} is the particle specific heat and T_p is the particular phase temperature. If the suspension is dilute, then the heat-transfer rate $\dot{q}_p$ can be expressed as[10]

$$\dot{q}_p = -n\dot{Q}_p, \tag{7}$$

where $\dot{Q}_p$ is the heat-transfer rate to a single particle. For the heat-transfer rate to a single gold nanoparticle, we will make use of the low frequency expression[20]

$$\dot{Q}_p = 4\pi r_0 k_f (T_p - T_f), \tag{8}$$

with r_0 nanoparticle radius and k_f gas thermal conductivity. The thermal conductivity k_f is in general a function of temperature, but in linear approximation it can be considered constant. For a dilute suspension, the distributed heat source $\dot{q}_{\text{laser}}$ (i.e., the laser source) can be written as[4]

$$\dot{q}_{\text{laser}} = n C_{\text{abs}} I, \tag{9}$$

where C_{abs} is the absorption cross-section for a single nanoparticle and I is the laser intensity. When the particles are much smaller than the wavelength of light, only the dipole contribution is important, and the absorption cross-section C_{abs} has a particularly simple expression,[5,27]

$$C_{\text{abs}} = 18\pi v_p \frac{(\varepsilon_m)^{3/2}}{\lambda} \frac{\varepsilon''}{(\varepsilon' + 2\varepsilon_m)^2 + \varepsilon''^2}, \tag{10}$$

where λ denotes the light wavelength, $\varepsilon = \varepsilon' + i\varepsilon''$ represents the complex dielectric function of the gold nanoparticle, and ε_m is the dielectric constant of the medium. Equation (10) has been derived from Mie theory and it is valid for small spherical particles.

Two-phase models are usually assumed to be unbounded. However, in our model the suspension has boundaries represented by the gas-liquid interface. A simplified boundary condition can be obtained by assuming that the bubble is spherical and the velocity u at the bubble wall is given by $u = \dot{R}$, where $R(t)$ is the instantaneous bubble radius and $\dot{R}$ is the velocity of the bubble wall.[28,29] Furthermore, when the compressibility of the liquid is neglected, the motion of the bubble wall can be described by the Rayleigh–Plesset equation,[30]

$$\rho_w \left(R\ddot{R} + \frac{3}{2}\dot{R}^2 \right) = p - p_\infty - \frac{4\mu_w \dot{R}}{R} - \frac{2\sigma_w}{R}, \tag{11}$$

with ρ_w the density of the liquid surrounding the bubble, p the pressure in the bubble-nanoparticle system, and p_∞ the undisturbed pressure in the liquid. The liquid surface tension is denoted by σ_w and the liquid viscosity by μ_w. In linear approximation, which is assumed in this study, only the linearized expression for Eq. (11) needs to be considered.

Due to the large numbers of variables involved, a partial list of variables and constants is presented in Nomenclature.

B. Linear equations for the gas-nanoparticle system

The confined two-phase model equations in Sec. II A have been solved in linear approximation. Substituting Eq. (3) into Eq. (2), we obtain the combined continuity equation:

$$\frac{D}{Dt}\rho_f + \frac{\rho_f}{1 - \phi_v} \nabla \cdot \boldsymbol{u} = 0. \tag{12}$$

Multiplying Eq. (12) for $c_{pf} T_f$ and employing the relation

$$c_{pf}\rho_f T_f = \frac{\gamma}{\gamma - 1} p \tag{13}$$

valid for an ideal gas with $\gamma = c_{pf}/c_{vf}$ ratio of specific heats, Eq. (12) can be written as

$$\rho_f c_{pf} \frac{D\rho_f}{Dt} + \frac{\gamma p}{\gamma - 1} \frac{1}{1 - \phi_v} \nabla \cdot \boldsymbol{u} = 0. \tag{14}$$

The energy equation for the fluid phase can be simplified using the ideal gas relation $\beta_f T_f = 1$, as

$$\rho_f c_{pf} \frac{D}{Dt} T_f - \dot{p} = \frac{\dot{q}_p}{1 - \phi_v}. \tag{15}$$

The sum of Eqs. (14) and (15) after using Eq. (13) yields the expression

$$\nabla \cdot \boldsymbol{u} = \frac{1}{r^2} \frac{\partial}{\partial r}(r^2 u) = \frac{(\gamma - 1)(1 - \phi_v)}{\gamma p} \left(\frac{\dot{q}_p}{1 - \phi_v} - \frac{\dot{p}}{\gamma - 1} \right). \tag{16}$$

The integration of Eq. (16) gives

$$u = -\frac{r}{3}\frac{(1-\phi_v)}{\gamma p}\dot{p} + \frac{r}{3}\frac{\gamma-1}{\gamma p}\dot{q}_p. \tag{17}$$

With the aid of the boundary condition at the bubble wall

$$u(R,t) = \dot{R}, \tag{18}$$

Eq. (17) can be turned into a differential equation for p as

$$\dot{R} = -\frac{1}{3}\frac{R\dot{p}}{\gamma p} + \frac{1}{3}\frac{R(\gamma-1)}{\gamma p}4\pi r_0 n k_f (T_p - T_f). \tag{19}$$

In Eq. (19) we made the approximation $1-\phi_v \approx 1$ valid for a diluted suspension. In Eqs. (17)–(19), we have neglected the spatial variation in the gas and particulate temperature. More accurate models should include such variation. This approximation simplifies the problem considerably, because it allows us to disregard the energy equation in the liquid.

The energy equations for both the gas phase and the particulate phase can be written in linear approximation, respectively, as

$$\rho_f c_{pf}\frac{dT_f}{dt} - 4\pi r_0 n k_f (T_p - T_f) = \dot{p}, \tag{20}$$

$$\rho_p c_p \frac{dT_p}{dt} + 4\pi r_0 n k_f (T_p - T_f) = n\sigma_{\mathrm{abs}} I. \tag{21}$$

For convenience, we will write Eqs. (19)–(21) in terms of the dimensionless variables $R=R_0R^*$, $p=p_0p^*$, $T_f=T_0T_f^*$, $T_p=T_0T_p^*$, where R_0 and p_0 are, respectively, the equilibrium radius and internal pressure of the bubble. R_0 and p_0 are related by Laplace's equation $R_0=2\sigma_w/(p_0-p_\infty)$. Furthermore, to first order one can write $R^*=1+X$ and $p^*=1+P$. By writing Eq. (19) in terms of those variable and omitting for simplicity the symbol $*$, we obtain

$$A(\gamma-1)T_f - A(\gamma-1)T_p + \dot{P} + 3\gamma\dot{X} = 0, \tag{22}$$

with A having the dimension of an inverse time defined by $A=(T_0/p_0)4\pi r_0 n k_f$.

Similarly for Eqs. (20) and (21), we have

$$(\Delta + AB)T_f - ABT_p - B\Delta P = 0, \tag{23}$$

$$(\Delta + AC)T_p - ACT_f = HC, \tag{24}$$

where $\Delta=d/dt$, and B, C, H are defined by $B=p_0/\rho_f c_{pf} T_0$, $C=p_0/\rho_p c_p \phi_v T_0$, and $H=n\sigma_{\mathrm{obs}}/p_0$. The coefficients B and C are dimensionless, and H has the dimension of an inverse time. The linearized form of the Rayleigh–Plesset equation may be written as[28]

$$\Delta^2 X - Z(P + WX - M\dot{X}) = 0, \tag{25}$$

with $Z=p_0/\rho_L R_0^2$, $W=2\sigma_w/R_0p_0$, and $M=4\mu_w/p_0$. The coefficient Z has the dimension of inverse time squared, W is dimensionless, and M has the dimension of time.

III. ANALYTICAL SOLUTIONS FOR A GAS-NANOPARTICLE STRUCTURE

After the linearization of the two-phase model described above, we obtain a system of differential equations with constant coefficients which in matrix notation reads as

$$\begin{pmatrix} \Delta+AB & -AB & -B\Delta & 0 \\ -AC & \Delta+AC & 0 & 0 \\ (\gamma-1)A & -(\gamma-1)A & \Delta & 3\gamma\Delta \\ 0 & 0 & Z & -\Delta^2 - ZM\Delta + ZW \end{pmatrix} \begin{pmatrix} T_f \\ T_p \\ P \\ X \end{pmatrix} = \begin{pmatrix} 0 \\ HC \\ 0 \\ 0 \end{pmatrix}. \tag{26}$$

When the right hand-side of Eq. (26) is zero, i.e., the laser pulse is off, a homogeneous set of equations is obtained. The mathematics associated with the solutions of the system of differential equations (26) is quite cumbersome and not entirely straightforward. Here, therefore, we provide an outline of a method for its solution.[31] The characteristic determinant of the system matrix in Eq. (26) can be written as

$$F = -\Delta^2(\Delta - m_1)(\Delta - m_2)(\Delta - m_3). \tag{27}$$

The order of the determinant F in Eq. (27) is 5, which is equal to the number of independent constants in the general solutions of the system of equations (26). Besides the eigenvalue $\Delta = 0$, Eq. (27) has the eigenvalues m_1, m_2, and m_3, which is the complex conjugate of m_2. Unfortunately, the equations for m_1 and m_2 are rather awkward and they can be evaluated only numerically. They can be written as

$$m_1 = x_1 - s/3, \quad m_2 = x_2 - s/3, \quad m_3 = m_2^* = x_3 - s/3, \tag{28}$$

where

$$x_1 = V_1 - V_2,$$

$$x_2 = -\frac{V_1 - V_2}{2} + i\sqrt{3}\frac{(V_1 + V_2)}{2},$$

$$x_3 = -\frac{V_1 - V_2}{2} - i\sqrt{3}\frac{(V_1 + V_2)}{2}, \tag{29}$$

and V_1 and V_2 are defined by

$$V_1 = \sqrt[3]{-\frac{b}{2} + \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}},$$

$$V_2 = \sqrt[3]{\frac{b}{2} + \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}}, \quad \frac{b^2}{4} + \frac{a^3}{27} > 0, \tag{30}$$

with $a=-\frac{1}{3}s^2+q$, $b=\frac{1}{27}[2s^3-9sq+27r]$, and s, q, and r are given by

$$s = AB\gamma + AC + ZM,$$

$$q = 3\gamma Z + ZMAB\gamma + ZMAC - ZW,$$

$$r = ZA[3\gamma C + 3\gamma B - \gamma WB - WC]. \tag{31}$$

For the diagonalization of the system of Eq. (26), we notice that an inhomogeneous system of $n \times n$ linear equations with constant coefficients can be diagonalized as

$$\frac{F}{\Gamma_i} y_i = \frac{G_{1i}}{\Gamma_i} f_1 + \frac{G_{2i}}{\Gamma_i} f_2 + \frac{G_{3i}}{\Gamma_i} f_3 + \cdots + \frac{G_{ni}}{\Gamma_i} f_n, \tag{32}$$

where $y_i (i=1,\ldots,n)$ are the solutions, F is the characteristic determinant of the system, G_{ij} are the co-factor of each element of the characteristic determinant, Γ_i is the highest common factor for $G_{li},\ldots,G_{ni}$, and f_i denotes the inhomogeneous part of the system of equations. For our system, we have $f_i=(0,HC,0,0)$ and $\Gamma_1=\Gamma_2=\Gamma_3=\Gamma_4=\Delta$. Hence, Eq. (32) reduces to

$$\frac{F}{\Delta} T_f = \frac{G_{21}}{\Delta} HC, \quad \frac{F}{\Delta} T_p = \frac{G_{22}}{\Delta} HC,$$

$$\frac{F}{\Delta} P = \frac{G_{23}}{\Delta} HC, \quad \frac{F}{\Delta} X = \frac{G_{24}}{\Delta} HC, \tag{33}$$

which translates into an uncoupled system of differential equations. By substituting the appropriate values for F and G_{ij} in Eq. (33), we obtain a system of differential equations that are easily solved:

$$\Delta[\Delta^3 + p\Delta^2 + q\Delta + r]T_f = \chi_1,$$

$$\Delta[\Delta^3 + p\Delta^2 + q\Delta + r]T_p = \chi_1,$$

$$\Delta[\Delta^3 + p\Delta^2 + q\Delta + r]P = -W\chi_4,$$

$$\Delta[\Delta^3 + p\Delta^2 + q\Delta + r]X = \chi_4, \tag{34}$$

with

$$\chi_1 = AB\gamma(3Z - ZW)HC,$$

$$\chi_4 = ZAHC(\gamma - 1). \tag{35}$$

The solutions of the system (34) can be expressed in terms of elementary functions as

$$T_f = a_0 + a_1 e^{m_1 t} + e^{\alpha t}[a_2 \cos(\beta t) + a_3 \sin(\beta t)] - \frac{\chi_1}{m_1(\alpha^2+\beta^2)} t,$$

$$T_p = b_0 + b_1 e^{m_1 t} + e^{\alpha t}[b_2 \cos(\beta t) + b_3 \sin(\beta t)] - \frac{\chi_1}{m_1(\alpha^2+\beta^2)} t,$$

$$P = c_0 + c_1 e^{m_1 t} + e^{\alpha t}[c_2 \cos(\beta t) + c_3 \sin(\beta t)] + \frac{W\chi_4}{m_1(\alpha^2+\beta^2)} t,$$

$$X = d_0 + d_1 e^{m_1 t} + e^{\alpha t}[d_2 \cos(\beta t) + d_3 \sin(\beta t)] - \frac{\chi_4}{m_1(\alpha^2+\beta^2)} t, \tag{36}$$

where α and β denotes the real and imaginary parts of the root m_2. The parameter β can be interpreted as the angular frequency of oscillations for the two-phase model.

The solution of the homogenous system is obtained by setting in $\chi_2=\chi_1=0$ in Eq. (36). Although the set of equations (36) fully determines T_f, T_p, P, X, yet, considered as a system, it is not necessarily equivalent to original system. The set of arbitrary constants in the solution of Eq. (36) greatly exceeds the order of our system that is 5. The exceeding constants can be eliminated by substituting Eq. (36) into the original system of differential equations, given by Eq. (26). By doing so, one obtains a system of 16 equations that are compatible. By solving this system of compatible equations, one obtains the general solution of Eq. (26). In our solution, the independent constants are b_0, c_1, c_2, c_3, and d_0. The remaining constants are rather complicated expressions of those five independent constants and they can be evaluated only numerically. They are given by

$$a_0 = b_0 + \frac{HCZ(3\gamma - W)}{m_1(\alpha^2+\beta^2)}, \quad c_0 = -\left[Wd_0 + \frac{MZAHC(\gamma-1)}{m_1(\alpha^2+\beta^2)}\right], \tag{37}$$

$$a_1 = B\left[\frac{\gamma - 1 + \psi}{\gamma - 1}\right] c_1, \quad b_1 = \frac{(m_1 + AB)\psi + AB(\gamma-1)}{\gamma - 1} c_1, \tag{38}$$

with ψ given by

$$\psi = 1 + \frac{3\gamma Z}{m_1^2 + ZMm_1 - ZW}. \tag{39}$$

The constant d_1 can be expressed in terms of c_1, and the constants d_2 and d_3 in terms of c_2 and c_3 as

$$d_1 = \frac{Z}{m_1^2 + ZMm_1 - ZW} c_1, \quad d_2 = \frac{-Z}{X_1^2 + X_2^2}[X_1 c_2 + X_2 c_3],$$

$$d_3 = \frac{Z}{X_1^2 + X_2^2}[X_1 c_2 - X_2 c_3], \tag{40}$$

with

$$X_1 = ZW - \alpha^2 + \beta^2 - ZM\alpha,$$

$$X_2 = 2\alpha\beta + ZM\beta. \tag{41}$$

The remaining coefficients a_2, a_3, b_2, b_3 are given, respectively, by

$$a_2 = \frac{\gamma B}{(\gamma-1)}\left\{\left[1 - \frac{3ZX_1}{X_1^2 + X_2^2}\right] c_2 - \frac{3ZX_2}{X_1^2 + X_2^2} c_3\right\},$$

$$a_3 = \frac{\gamma B}{(\gamma-1)}\left\{\frac{3ZX_2}{X_1^2 + X_2^2} c_2 + \left(1 - \frac{3ZX_1}{X_1^2 + X_2^2}\right) c_3\right\}, \tag{42}$$

$$b_2 = \frac{1}{(\gamma-1)A}\left\{\left[\alpha + AB\gamma - \frac{3\gamma Z(\alpha+AB)}{X_1^2+X_2^2}X_1 + \frac{3\gamma\beta Z}{X_1^2+X_2^2}X_2\right]c_2 + \left[\beta - \frac{3\gamma Z(\alpha+AB)}{X_1^2+X_2^2}X_2 - \frac{3\gamma\beta Z}{X_1^2+X_2^2}X_1\right]c_3\right\}, \quad (43)$$

$$b_3 = \frac{1}{(\gamma-1)A}\left\{\left[-\beta + \frac{3\gamma\beta Z}{X_1^2+X_2^2}X_1 + \frac{3\gamma Z(\alpha+AB)}{X_1^2+X_2^2}X_2\right]c_2 + \left[\alpha + AB\gamma - \frac{3\gamma Z(\alpha+AB)}{X_1^2+X_2^2}X_1 + \frac{3\gamma\beta Z}{X_1^2+X_2^2}X_2\right]c_3\right\}. \quad (44)$$

IV. ANALYTICAL SOLUTIONS FOR A LIQUID-NANOPARTICLE STRUCTURE

It is interesting to compare our gas-nanoparticle system to a similar structure having the gas replaced by a liquid. This system may represent, for example, a droplet (liposome) containing gold nanoparticles dispersed in an aqueous solution. Our linear model, described in Sec. II B, is based on the fact that the fluid inside the bubble is an ideal gas and the equation of state and other well known relationships valid for ideal gases are employed. If the bubble contains a liquid, however, those equations can no longer be applied. Because we are working in linear approximation, we can, however, make use of a linearized equation of state valid for the fluid as follows. The combined continuity equation given by Eq. (12) can be written in linearized form as

$$(1-\phi_{v0})\frac{\partial}{\partial t}\rho_f' + \rho_{f0}\nabla\cdot\boldsymbol{u} = 0, \quad (45)$$

where ρ_f', ρ_{f0} are related by the linear relation $\rho_f = \rho_{f0} + \rho_f'$ and ϕ_{v0} is the undisturbed particle volume concentration. The linearized form of the energy equation for the fluid, Eq. (15), is given by

$$\rho_{f0}c_{pf}\frac{\partial T_f'}{\partial t} - \dot{p}_f' = \frac{\dot{q}_p}{1-\phi_{v0}}, \quad (46)$$

where $T_f = T_0 + T_f'$ and $p_f = p_0 + p_f'$, and $1-\phi_{v0} \approx 1$. The fluid equation of state in linear form is[20]

$$\rho_f' = \frac{p_f'}{c_{Tf}^2} - \rho_{f0}\beta_f T_f', \quad (47)$$

where c_{Tf} is the isothermal speed of sound if the fluid and β_f is the coefficient of thermal expansion for the fluid. After multiplying Eq. (45) with c_{pf}/β_f, substituting Eq. (47) into Eq. (45) and adding Eqs. (47) and (45), we obtain

$$\nabla\cdot\boldsymbol{u} = \gamma_1'\dot{q}_p - (\gamma_2' - \gamma_1')\dot{p}_f', \quad (48)$$

with $\gamma_1' = \beta_f/\rho_{f0}c_{pf}$, $\gamma_2' = 1/\rho_{f0}c_{Tf}^2$. After integration and using the same notation as in the previous sections, Eq. (48) can be written as

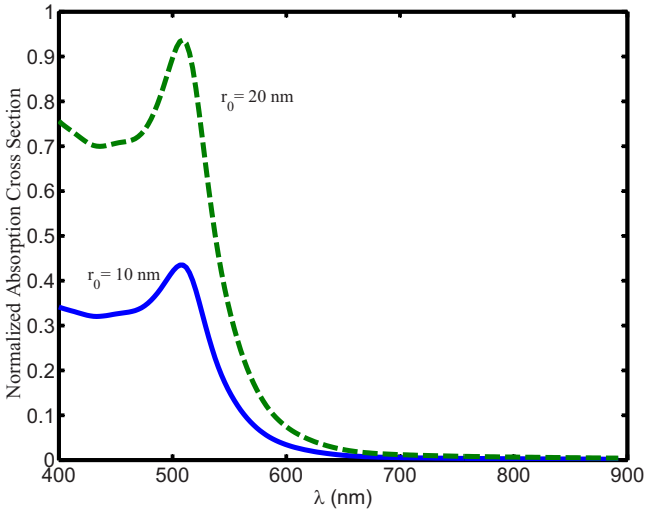


FIG. 1. (Color online) Dimensionless absorption cross-section as a function of the light wavelength for a nanoparticle of radii 10 and 20 nm.

$$(\gamma_2\gamma_3 - 1)AT_f - (\gamma_2\gamma_3 - 1)AT_p + \Delta P + 3\gamma_3\Delta X = 0, \quad (49)$$

with γ_3, γ_2 dimensionless variables defined by

$$\gamma_3 = \frac{1}{\gamma_2 - \gamma_1}, \quad \gamma_1 = \frac{p_0\beta_f}{\rho_f c_{pf}}, \quad \gamma_2 = \frac{p_0}{\rho_f c_{Tf}^2}. \quad (50)$$

The other equations for the liquid-nanoparticle droplet are the same as for the gas-nanoparticle system. Therefore, the system can be solved as described in Sec. III, with the third equation in the system of equations (26) replaced by Eq. (49). Formally, the eigenvalue expressions, Eqs. (27)–(30) and the eigenfunctions, Eq. (36), remain the same but some of the coefficients in those equations change. The expressions which are different are given in Appendix A.

V. NUMERICAL EXAMPLES

In this section, the numerical predictions of the mathematical model described above will be illustrated. We have considered here regions of the parameter space where the linear approximation can be applied.

A. Gold nanoparticle absorption cross-section

For the purpose of illustration, we have plotted the dimensionless absorption cross-section $K_{abs} = C_{abs}/\pi r_0^2$ with C_{abs} given by Eq. (10). We assume that the environment surrounding the nanoparticle is gas and therefore its dielectric constant $\varepsilon_m \cong 1$. The data used for the dielectric function of gold were taken from Ref. 32, which provides the dielectric function of bulk gold in the optical range. Strictly speaking the dielectric function of a bulk metal cannot be used to describe the optical properties of nanoparticles, because of quantum mechanical and surface effects. However, for gold nanoparticles with radius 10 nm and larger, the bulk dielectric function gives a reasonably good agreement between the experimental spectra and the Mie theory.[33] In Fig. 1, K_{abs} is plotted as a function of the light wavelength for two different nanoparticle radii $r_0 = 10$ nm and 20 nm. As may be seen, the dimensionless absorption cross-section shows a maximum at a wavelength of about $\lambda = 508$ nm and more light is ab-

TABLE I. Input parameters (Refs. 34 and 35)

Parameter	Symbol	Unit	
Water density	ρ_w	kg/m^3	0.998×10^3
Water viscosity	μ_w	N/m^2 s	0.923×10^{-3}
Thermal conductivity of water	k_f	W/m K	0.5984
Thermal expansion coefficient of water	β_f	1/K	2.10×10^{-4}
Surface tension	σ_w	N/m	72.82×10^{-3}
Isothermal speed of sound in water	c_{fT}	m/s	1481.8
Water specific heat at constant pressure	c_{pf}	J/kg K	4182
Ambient temperature	T_0	K	298
Ambient pressure	p_∞	bar	1.0
Thermal conductivity C_3F_8	k_f	W/m K	0.011 35
Density of gold	ρ_g	kg/m^3	1.93×10^4
Specific heat of gold	c_{pp}	J/kg K	1.29×10^2
Ratio of specific heats for an ideal gas	γ		1.4
C_3F_8 specific heat at constant pressure	c_{pf}	J/kg K	0.794×10^3

sorbed by the larger size nanoparticle than the smaller one. The numerical values[34,35] used to plot Fig. 1 and the remaining figures are given in Table I.

B. Bubble-nanoparticle system eigenvalues

The two-phase model predicts a real eigenvalue m_1 and a complex eigenvalue m_2. The m_1 eigenvalue corresponds to an exponentially decaying motion and the complex one to an exponentially decaying oscillatory motion. In Fig. 2 we have plotted m_1 in units of (1/μs) as a function of the gas bubble ambient radius R_0 for nanoparticle radii $r_0=10$ nm and 20 nm at different particle volume concentration: $\phi_v=10^{-6}$, 10^{-5}, 10^{-4}. We have assumed that the bubble is made of perfluorocarbon (C_3F_8). Similar results are obtained for air. As may be seen in the Fig. 2, m_1 is a slowly varying function of R_0 and remains negative in the region of the parameter space considered. m_1 becomes more negative for increasing particle volume concentration and becomes less negative for increasing nanoparticle size.

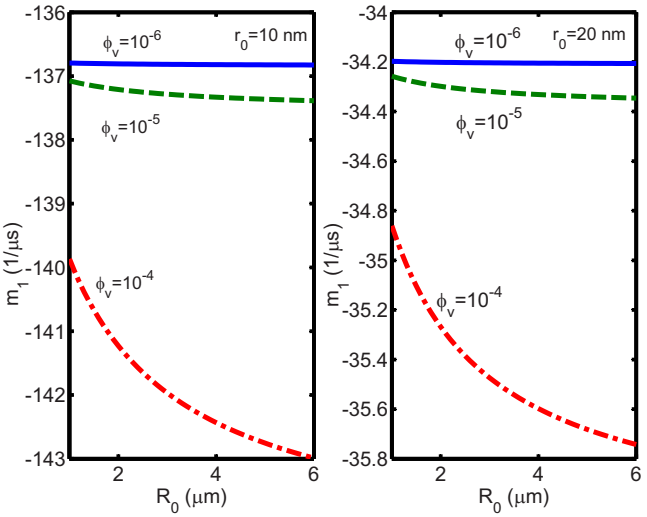


FIG. 2. (Color online) Eigenvalue m_1 as a function of the bubble equilibrium radius for a nanoparticle of radius 10 nm and a nanoparticle of radius 20 nm.

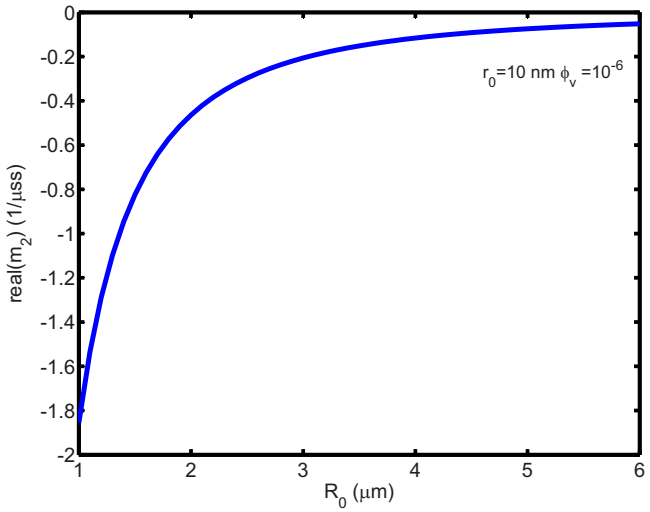


FIG. 3. (Color online) Real part of the eigenvalue m_2 as a function of the bubble equilibrium radius for a nanoparticle of radius 10 nm.

In Fig. 3, the real part of $m_2(\alpha)$ has been plotted as a function of the bubble radius for $\phi_v=10^{-6}$ and $r_0=10$ nm. Similar plots have been obtained for $\phi_v=10^{-5}-10^{-4}$, and $r_0=20$ nm, and therefore have not been shown. As m_1, α is negative but is it much smaller than m_1 in absolute value. Therefore, the oscillatory motion decays more slowly than the non-oscillatory one. Furthermore, α is much less sensitive to the variation in r_0 and ϕ_v than m_1, and it is an increasing function of R_0.

In Fig. 4, the imaginary part of $m_2/2\pi$ ($\beta/(2\pi)$) has been plotted in megahertz for $r_0=10$ nm and $\phi_v=10^{-6}$. $\beta/2\pi$ remains essentially the same in the parameter space checked, i.e., $\phi_v=10^{-5}$, 10^{-4}, and $r_0=10-20$ nm. For the purpose of comparison, the angular resonance frequency ω_0 of a gas bubble in an unbounded liquid (water)[30]

$$f_0=\frac{1}{2\pi}\sqrt{\frac{3\kappa p_0}{\rho_w R_0^2}-\frac{2\sigma_L}{\rho_w R_0^3}} \tag{51}$$

has also been plotted in Fig. 4. For isothermal behavior, $\kappa\cong1$. As for f_0, β decreases for increasing bubble radius; how-

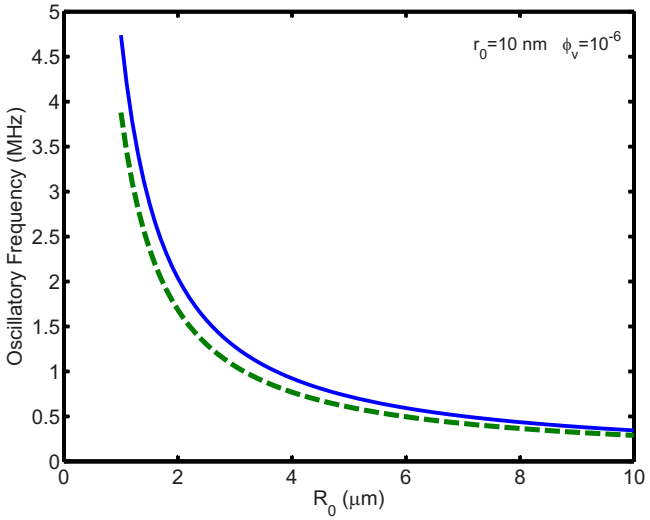


FIG. 4. (Color online) Oscillatory frequency of the bubble-nanoparticle system and resonance frequency of a free gas bubble as a function of the bubble equilibrium radius.

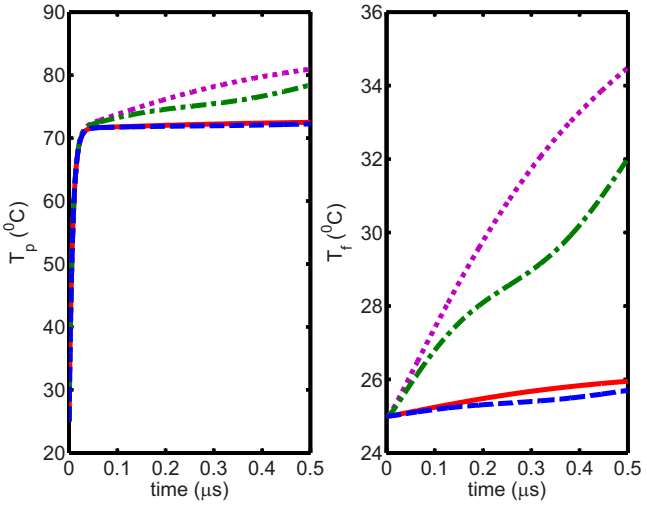


FIG. 5. (Color online) Particulate and gas temperatures as a function of time during the laser exposure for different bubble radii and particle volume concentrations. The dotted (···) line is for $R_0=4$ μm and $\phi_v=10^{-5}$, the solid line (—) for $R_0=4$ μm and $\phi_v=10^{-6}$, the dashed-dotted (-·-·-) line for $R_0=2$ μm and $\phi_v=10^{-5}$, and the dashed line (---) for $R_0=2$ μm and $\phi_v=10^{-6}$. The pulse duration is $t_L=0.5$ μs and the laser intensity $I_L=0.5\times10^9$ W/m^2.

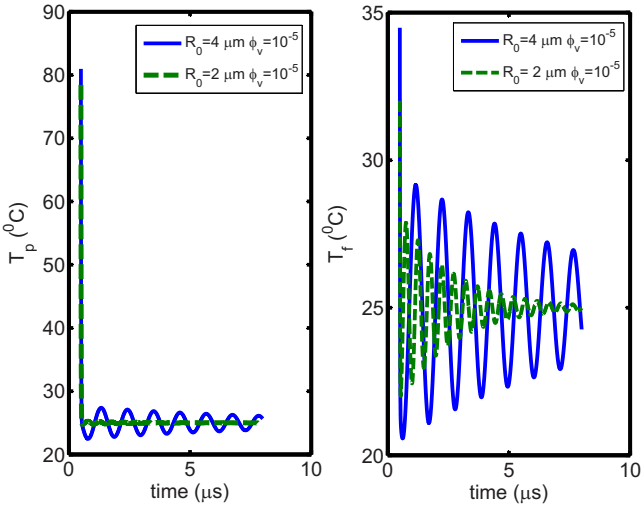


FIG. 6. (Color online) Particulate and gas temperatures as a function of time after the end of the laser pulse for $t_L=0.5$ μs and laser intensity $I_L=0.5\times10^9$ W/m^2.

ever f_0 is smaller than $\beta/2\pi$ in the region of the parameter space illustrated in the plot.

C. Temperature for the bubble-nanoparticle system

The temperature of the particulate phase and the gas phase are plotted in Fig. 5 as a function of time for two different bubble radii $R_0=2$, 4 μm, nanoparticle radius $r_0=10$ nm, and particle volume concentration $\phi_v=10^{-6}$, 10^{-5}. The laser exposure conditions are $\lambda = 508$ nm, $I_L=0.5\times10^9$ W/m^2, and pulse duration $t_L=0.5$ μm. As may be seen in Fig. 5, the increase in the gas temperature, T_f, remains below 26 °C for $\phi_v=10^{-6}$ and increases of a few degrees at the end of the laser pulse for the higher particle volume concentration $\phi_v=10^{-5}$. The temperature of the particulate phase, T_p, instead increases very quickly to about 70 °C. Bubble size and particle volume concentration does not significantly affect the particulate temperature in the region of the parameter space considered. In Fig. 6, the gas temperature and the particulate temperature after the laser pulse have been plotted as a function of time for the same laser intensity as in Fig. 5. The gas temperature undergoes sinusoidal decaying oscillations and the oscillations are stronger for the $R_0=4$ μm, $\phi_v=10^{-5}$ bubble than for the $R_0=2$ μm, $\phi_v=10^{-6}$. The particulate temperature instead decays very quickly to the ambient temperature without significant oscillations. During the laser pulse, the bubble expands and its oscillations occur mainly immediately after the laser pulse. The temperature reached by the particulate phase and the gas phase at the end of the laser pulse are therefore a strong indicator of the strength of the subsequent oscillations.

D. Bubble gas pressure and oscillations

The dimensionless pressure p/p_0 and the dimensionless bubble radius R/R_0 have been plotted in Fig. 7 for the same exposure conditions, bubble radii, and particle volume concentration as in Fig. 5. As may be expected, the pressure oscillations are larger for the larger size bubble $R_0=4$ μm and higher particle volume concentration than all the other cases considered. The dimensionless bubble radius R/R_0 is an increasing function of time and the increase is stronger for the larger particle volume concentration. In Fig. 8, we have plotted the dimensionless pressure p/p_0 and the dimensionless bubble radius R/R_0 as a function of time after the laser exposure. Both variables undergo sinusoidal decaying oscillations with the stronger oscillations occurring for the larger size bubble and the higher particle volume concentration. The values of the dimensionless pressure p/p_0 and the dimensionless bubble radius R/R_0 at the end of the laser pulse have been plotted as a function of pulse duration in Fig. 9. As may be seen, the increase in absolute value of the pressure and bubble radius at the end of the pulse is modest for

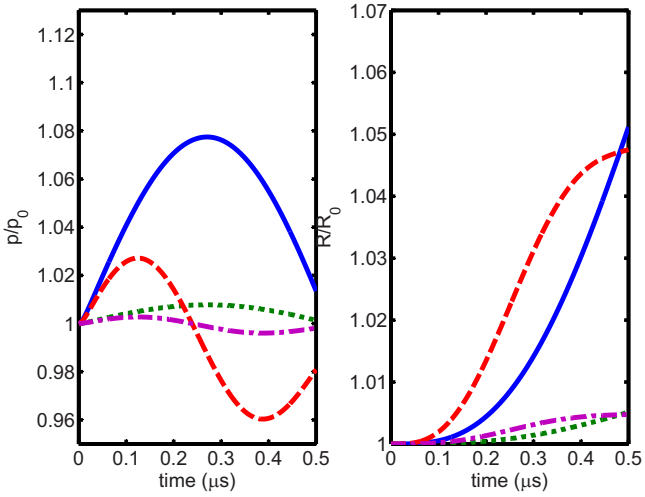


FIG. 7. (Color online) Dimensionless gas pressure and dimensionless bubble radius as a function of time during the laser exposure for different bubble radii and particle volume concentrations. The dotted line (···) is for $R_0=4$ μm and $\phi_v=10^{-6}$, the solid line (—) for $R_0=4$ μm and $\phi_v=10^{-5}$, the dashed-dotted line (-·-·-) for $R_0=2$ μm and $\phi_v=10^{-6}$, and the dashed line (---) for $R_0=2$ μm and $\phi_v=10^{-5}$. The pulse duration is $t_L=0.5$ μs and the laser intensity $I_L=0.5\times10^9$ W/m^2.

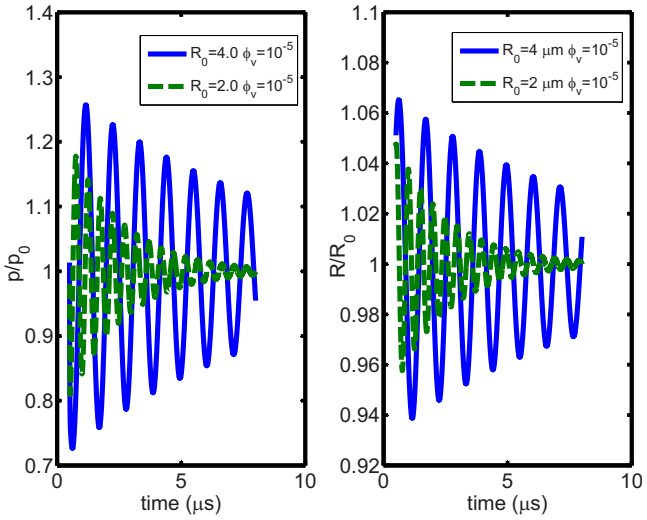


FIG. 8. (Color online) Dimensionless gas pressure and dimensionless bubble radius as a function of time after the end of the laser pulse for $t_L=0.5$ μs and laser intensity $I_L=0.5\times10^9$ W/m^2.

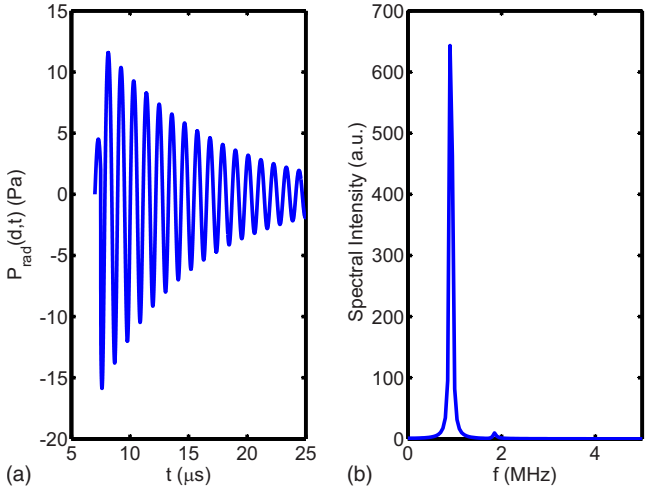


FIG. 10. (Color online) (a) Radiated pressure as a function of time for a bubble containing gold nanoparticles. (b) Spectral intensity of the radiated pressure occurring after the end of the laser pulse for a bubble-nanoparticle system of radius 4 μm. The pulse duration is $t_L=0.5$ μs and the laser intensity $I_L=0.5\times10^9$ W/m^2.

the lowest intensity considered 0.1×10^9 W/m^2, but it becomes significant for the highest intensity and concentration considered.

E. Transduction efficiency gain due to gas bubble

Figure 4 shows the frequency of oscillation $\beta/2\pi$ for our bubble-nanoparticle structure as a function of the bubble radius R_0. For example, at $R_0=4$ μm, the linear resonance frequency is $f_0=\beta/(2\pi)=0.92$ MHz. One can estimate the far-field radiated pressure at a distance $d\gg R_0$ as[36,37]

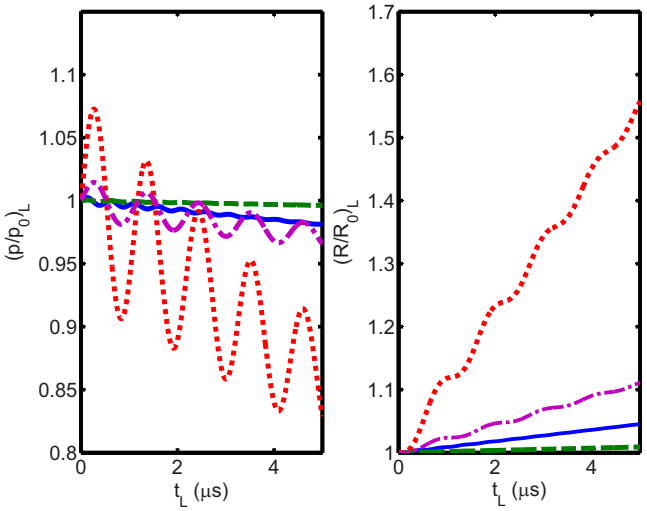


FIG. 9. (Color online) Gas pressure and bubble radius values reached at the end of the laser pulse as a function of the pulse duration t_L for different bubble radii, particle volume concentrations, and exposure conditions. The dotted line is for $R_0=4$ μm, $\phi_v=10^{-5}$, and $I_L=0.5\times10^9$ W/m^2; the solid line for $R_0=2$ μm, $\phi_v=10^{-6}$ and, $I_L=0.5\times10^9$ W/m^2; the dashed-dotted line or $R_0=4$ μm, $\phi_v=10^{-5}$, and $I_L=0.1\times10^9$ W/m^2; and the dashed line for $R_0=2$ μm, $\phi_v=10^{-6}$, and $I_L=0.1\times10^9$ W/m^2.

$$P_{\rm rad}(d,t)=\frac{\rho_w}{4\pi d}\frac{d^2V}{dt^2}=\rho_w\frac{R}{d}(2\dot{R}^2(t')+R(t')\ddot{R}(t')), \quad (52)$$

where t' is the retarded time $t'=t-r/c_w$, with c_w speed of sound in the water. Denoting the fast Fourier transform of $P_{\rm rad}(d,t)$ as $P_{\rm rad}(d,f)$, the *spectral intensity* of the emitted sound at the distance d can be defined as $I_0(d,f)=|P_{\rm rad}(d,f)|^2/2\rho_w c_w$. Every frequency component $I_0(f)$ is attenuated as the wave travels through the liquid. The linear attenuation of the wave in water can be expressed as

$$I(d,f)=I_0(d,f)\exp(-\alpha(f)d), \quad (53)$$

where $\alpha(f)$ is the attenuation coefficient which in water is[38] $\alpha=b(f/c_w)^2$ with $b=5.7\times10^{-6}$ cm.

In Fig. 10, the radiated pressure $P_{\rm rad}$, Eq. (52), at a distance $d=1$ cm has been plotted as a function of time during and after the laser pulse. In the same plot, the attenuated spectral intensity, Eq. (53), associated with the pressure radiated by the bubble at the end of the laser pulse has also been plotted. The plots have been obtained assuming $R_0=4$ μm, $t_L=0.5$ μs, $I_L=0.5\times10^9$ W/m^2, $\phi_v=10^{-5}$, $r_0=10$ nm. As may be expected, the radiated intensity peaks at the natural frequency of oscillations for the bubble-nanoparticle system. The attenuation coefficient $\exp(-\alpha(f)d)$ is negligible in the low megahertz range and can be neglected.

For the case of a liquid-nanoparticle system, the droplet undergoes radial oscillations, which are much smaller than for the gas-nanoparticle system with R/R_0 of the order 10^{-6}. The natural frequency of oscillation is, however, much larger than for the bubble structure. For example, for a radius of $R_0=4$ μm, the natural frequency of oscillation is $f_0=\beta/(2\pi)=108$ MHz. This value has been obtained by solving numerically Eq. (28). As may be seen in Fig. 11, the spectral intensity of the attenuated radiated pressure has a maximum at the same frequency. Figure 11 has been obtained for the same exposure conditions as Fig. 10. For simplicity, we have assumed that the liquid inside the droplet has

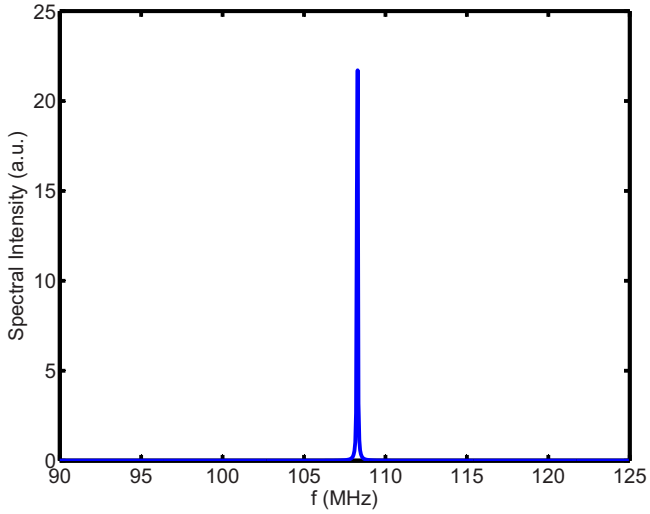


FIG. 11. (Color online) Attenuated spectral intensity for the radiated pressure occurring after the end of the laser pulse for a droplet-nanoparticle system of radius 4 μm. The pulse duration is $t_L=0.5$ μs and the laser intensity $I_L=0.5\times10^9$ W/m^2.

water properties. For the purpose of comparison, in Fig. 12, we have plotted the radiated pressure, Eq. (52), and the attenuated radiated pressure $P_{\text{att}}(d,f)=P_{\text{rad}}\exp(-\alpha(f)d/2)$. Although the radial oscillations of the droplet is small, the radiated pressure is comparable with the corresponding pressure for the bubble structure because of the acceleration term in Eq. (52). However, the pressure wave is much more attenuated owing to the much higher natural frequency of oscillations. Therefore, although the laser energy absorption is actually less effective in a gas than in liquid, there is a clear advantage of putting the gold nanoparticles in the gas, which results from having bubble resonant oscillations at the megahertz range and less attenuation for the radiated pressure.

VI. DISCUSSION AND CONCLUSIONS

In medical applications, preformed gas bubbles are employed to enhance ultrasound scattering from blood in order to facilitate studies of blood-filled cavities and blood flow. Photoacoustic excitation is an alternative imaging strategy that promises enhanced contrast over ultrasound imaging and enhanced resolution and depth of penetration over optical imaging. Two major issues with photoacoustic imaging are the lack of coherent penetration of light into tissues, requiring highly efficient optical absorbers to compensate, and the inefficient nature of thermo-acoustic transduction at the megahertz frequency. This led us to the question of whether it might be possible to combine gold nanoparticles with microbubbles to give efficient optical absorption and ultrasonic generation specifically in the megahertz frequency range. While this idea seems reasonable from the outset, the practical implementation is challenging, justifying a more rigorous look at the possible advantages of this approach.

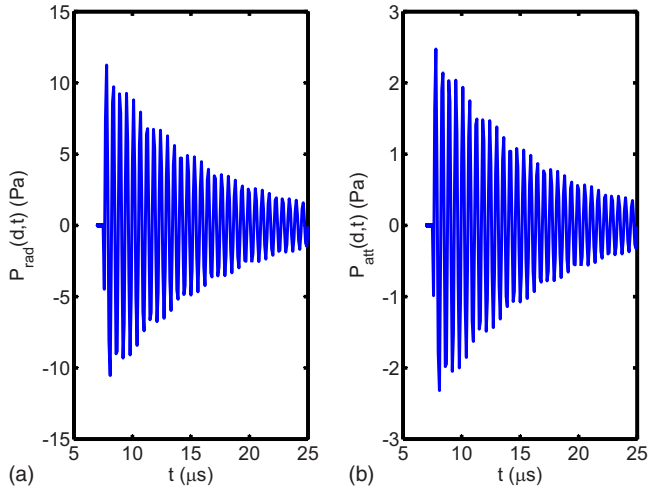


FIG. 12. (Color online) (a) Radiated pressure as a function of time for a droplet containing gold nanoparticles. (b) Attenuated radiated pressure as a function of time for a droplet containing GNPs after transmission through 1 cm of water. The exposure conditions are as in Fig. 11.

For simplicity, in this paper, we have considered spherical nanoparticles that have absorption spectra in the visible region, but in principle one could extend our analysis to nanorods and nanoshells that have absorption spectra in the near-infrared region, where optical transmission through tissue is better than for visible light.

This theoretical work is a first preliminary study on the possibility of inducing oscillations in a preformed microbubble containing spherical gold nanoparticles. Such a bubble has been analyzed in terms of a two-phase model. In general, two-phase models can be solved only numerically and in our case the problem is further complicated by the presence of boundaries represented by the bubble-liquid interface. Given the complexity of the subject, only a few aspects of the behavior of the microbubble-nanoparticle system have been addressed in this investigation. Our study focuses on its linear behavior. This leads to analytical expressions for the radial oscillations, the temperature, and the oscillation frequency of the system. This result has been obtained at the price of a number of approximations, the most notably of which are uniform pressure and temperature distribution in the bubble. As a result, these expressions are not rigorously exact. Nevertheless, when applied to situations of importance in medical applications, they offer useful insight into the planning and the analysis of experiments, and they provide a foundation upon which more realistic numerical models could be developed. The calculation illustrated here represents one of the few examples of a two-phase model with boundaries which can be solved analytically.

In conclusion, we have investigated the radial oscillations of a micrometer sized bubble containing spherical gold nanoparticles which is set into resonant oscillations by laser pulses of appropriate wavelength, duration, and intensity. In linear approximation and assuming a low-frequency expression for the heat exchange between the gas and the particles, we have derived analytical results for the bubble pressure, the gas, and the nanoparticle temperature, as well as the bubble radial oscillations. We have shown that the bubble undergoes sinusoidal decaying oscillations at frequencies which are in the low megahertz range and that, in contrast, a liquid-nanoparticle structure show sinusoidal decaying oscillations at frequencies which are about two orders of magnitude higher. This narrowband behavior is in contrast to that of more conventional photoacoustic contrast agents, which

have no resonance and whose frequency content is therefore determined by the laser pulse parameters and the thermal properties of the medium. Narrowband behavior means better matching to the receiving transducer and thus improved sensitivity. As well, less energy is lost in transmission due to tissue absorption at high frequencies. The motivation of this study lies in the possibility of using microbubbles containing light absorbing nanoparticles as contrast agents for novel photoacoustic applications, including molecular imaging.

ACKNOWLEDGMENT

This work was supported by the Methodist Hospital Research Institute.

NOMENCLATURE

A = Constant in the two-phase model defined in Eq. (22)
B = Constant in the two-phase model defined in Eq. (23)
c_{pf} = Gas specific heat at constant pressure
c_{pp} = Particle specific heat
c_{Tf} = Fluid isothermal speed of sound
C = Constant in the two-phase model defined in Eq. (24)
C_{abs} = Absorption cross-section
H = Constant in the two-phase model defined in Eq. (24)
I = Laser intensity
k_f = Gas thermal conductivity
K_{abs} = Dimensionless absorption cross-section
M = Constant in the linearized Rayleigh–Plesset equation (Eq. (25))
n = Number of particles per unit suspension volume
$m_{1,2,3}$ = Eigenvalues of the two-phase model Eq. (28)
R = Bubble radius
R_0 = Bubble equilibrium radius
r_0 = Nanoparticle radius
p_o = Equilibrium pressure on the gas bubble
p = Gas pressure
p_∞ = Undisturbed pressure in the liquid
P = Dimensionless gas pressure
$\dot{q}_p$ = Distributed particle heat-transfer rate per unit volume of suspension
$\dot{Q}_p$ = Heat-transfer rate to a single particle
$\dot{q}_{\text{laser}}$ = Distributed heat source
T_f = Gas temperature
T_p = Particulate temperature
T_0 = Undistributed (ambient) temperature
t_L = Laser pulse duration
$\mathbf{u}$ = Radial velocity
v_p = Particle volume
$V_{1,2}$ = Constants in the two-phase model defined in Eq. (30)
W = Constant in the linearized Rayleigh–Plesset equation [Eq. (25)]
X = Dimensionless bubble radius
Z = Constant in the linearized Rayleigh–Plesset equation [Eq. (25)]
α = Real part of the eigenvalue m_2
β = Imaginary part of the eigenvalue m_2
β_f = Fluid coefficient of thermal expansion
γ = Ratio of specific heats for the gas
$\gamma_{1,2,3}$ = Dimensionless coefficients defined in Eq. (50)
ε = Complex dielectric constant of gold
ε_m = Dielectric constant of the gas
ε' = Real part of the dielectric constant of gold
ε'' = Imaginary part of the dielectric constant of gold
λ = Light wavelength
μ_L = Liquid viscosity
ρ_L = Liquid density
ρ_f = Gas density
ρ_p = Density of the particle material
σ_f = Density of the gas phase
σ_p = Density of the particulate phase
σ_L = Liquid surface tension
ϕ_v = Particle volume concentration
ω_0 = Bubble free resonance frequency [Eq. (51)]

APPENDIX A: COEFFICIENTS FOR THE LIQUID-NANOPARTICLE SYSTEM

For a droplet containing gold nanoparticles, the analytical solutions described in Sec. III remain valid but some of the coefficients defined in that section are changed as follows. The coefficient s, q, and r defined by Eq. (31) are now given by

$$s = AB\gamma_2\gamma_3 + AC + ZM,$$

$$q = 3\gamma_3 Z + ZMAB\gamma_2\gamma_3 + ZMAC - ZW,$$

$$r = ZA[3\gamma_3 C + 3\gamma_3 B - \gamma_2\gamma_3 WB - WC]. \tag{A1}$$

The coefficients χ_1 and χ_2, Eq. (35), are now written as

$$\chi_1 = AB\gamma_3(3Z - ZW\gamma_2)HC,$$

$$\chi_4 = ZAHC(\gamma_2\gamma_3 - 1). \tag{A2}$$

The coefficients a_0, c_0, a_1, and b_1 are expressed by

$$a_0 = b_0 + \frac{HCZ(3\gamma_3 - W)}{m_1(\alpha^2 + \beta^2)},$$

$$c_0 = -\left[Wd_0 + \frac{MZAHC(\gamma_2\gamma_3 - 1)}{m_1(\alpha^2 + \beta^2)}\right], \tag{A3}$$

$$a_1 = B\left[\frac{\gamma_1\gamma_3 - 1 + \psi}{\gamma_2\gamma_3 - 1}\right]c_1,$$

$$b_1 = \frac{(m_1 + AB)\psi + AB(\gamma_2\gamma_3 - 1)}{\gamma_2\gamma_3 - 1}c_1, \tag{A4}$$

with ψ defined by

$$\psi = 1 + \frac{3\gamma_3 Z}{m_1^2 + ZMm_1 - ZW}. \tag{A5}$$

The remaining coefficients a_2, a_3, b_2, b_3 are given, respectively, by

$$a_2 = \frac{\gamma_3 B}{(\gamma_2\gamma_3 - 1)}\left\{\left[\gamma_2 - \frac{3ZX_1}{X_1^2 + X_2^2}\right]c_2 - \frac{3ZX_2}{X_1^2 + X_2^2}c_3\right\},$$

$$a_3 = \frac{\gamma_3 B}{(\gamma_2\gamma_3 - 1)}\left\{\frac{3ZX_2}{X_1^2 + X_2^2}c_2 + \left(\gamma_2 - \frac{3ZX_1}{X_1^2 + X_2^2}\right)c_3\right\}, \tag{A6}$$

$$b_2 = \frac{1}{(\gamma_2\gamma_3 - 1)A}\left\{\left[\alpha + AB\gamma_2\gamma_3 - \frac{3\gamma_3 Z(\alpha + AB)}{X_1^2 + X_2^2}X_1 + \frac{3\gamma_3\beta Z}{X_1^2 + X_2^2}X_2\right]c_2 + \left[\beta - \frac{3\gamma_3 Z(\alpha + AB)}{X_1^2 + X_2^2}X_2 - \frac{3\gamma_3\beta Z}{X_1^2 + X_2^2}X_1\right]c_3\right\}, \tag{A7}$$

$$b_3 = \frac{1}{(\gamma_2\gamma_3 - 1)A}\left\{\left[-\beta + \frac{3\gamma_3\beta Z}{X_1^2 + X_2^2}X_1 + \frac{3\gamma_3 Z(\alpha + AB)}{X_1^2 + X_2^2}X_2\right]c_2 + \left[\alpha + AB\gamma_2\gamma_3 - \frac{3\gamma_3 Z(\alpha + AB)}{X_1^2 + X_2^2}X_1 + \frac{3\gamma_3\beta Z}{X_1^2 + X_2^2}X_2\right]c_3\right\}, \tag{A8}$$

where X_1 and X_2 are defined in Eq. (41).

[1] T. G. Leighton, "From seas to surgeries, from babbling brooks to baby scans: The acoustics of gas bubbles in liquids," Int. J. Mod. Phys. B **18**, 3267–3314 (2004).
[2] H. Becher and P. N. Burns, *Handbook of Contrast Echocardiography* (Springer, Berlin, 2000).
[3] M. Xua and L. V. Wang, "Photoacoustic imaging in biomedicine," Rev. Sci. Instrum. **77**, 041101 (2006).
[4] U. Kreibig and M. Vollmer, *Optical Properties of Metal Clusters* (Springer-Verlag, Berlin, 1995).
[5] C. Bohren and D. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley-International, New York, 1983).
[6] M. Hu, J. Chen, Z. Y. Li, L. Au, G. V. Hartland, X. Li, M. Marquez, and Y. Xia, "Gold nanostructures: Engineering their plasmonic properties for biomedical applications," Chem. Soc. Rev. **35**, 1084–1094 (2006).
[7] P. K. Jain, X. Huang, I. H. El-Sayed, and M. A. El-Sayed, "Review of some interesting surface plasmon resonance-enhanced properties of noble metal nanoparticles and their applications to biosystems," Plasmonics **2**, 107–118 (2007).
[8] R. A. Sperling, P. R. Gil, F. Zhang, M. Zanella, and W. J. Parak, "Biological applications of gold nanoparticles," Chem. Soc. Rev. **37**, 1896–1906 (2008).
[9] B. Krasovitski, H. Kislev, and E. Kimmel, "Modeling photothermal and acoustical induced microbubble generation and growth," Ultrasonics **47**, 90–101 (2007).
[10] D. Lapotko, "Optical excitation and detection of vapor bubbles around plasmonic nanoparticles," Opt. Express **17**, 2538–2556 (2009).
[11] V. Kotaidis, C. Dahmen, G. von Plessen, F. Springer, and A. Plech, "Excitation of nanoscale vapor bubbles at the surface of gold nanoparticles in water," J. Chem. Phys. **124**, 184702 (2006).
[12] V. P. Zharov, E. N. Galitovskaya, C. Johnson, and T. Kelly, "Synergistic enhancement of selective nanophotothermolysis with gold nanoclusters: potential for cancer therapy," Lasers Surg. Med. **37**, 219–226 (2005).
[13] E. Y. Hleb, J. H. Hafner, J. N. Myers, E. Y. Hanna, B. C. Rostro, S. A. Zhdanok, and D. O. Lapotko, "LANTCET: Elimination of solid tumor cells with photothermal bubbles generated around clusters of gold nanoparticles," Nanomedicine **3**, 647–667 (2008).
[14] G. Wu, A. Mikhailovsky, H. A. Khant, C. Fu, W. Chiu, and J. A. Zasadzinski, "Remotely triggered liposome release by near-infrared light absorption via hollow gold nanoshells," J. Am. Chem. Soc. **130**, 8175–8177 (2008).
[15] E. Y. Hleb, Y. Hu, R. A. Drezek, J. H. Hafner, and D. O. Lapotko, "Photothermal bubbles as optical scattering probes for imaging living cells," Nanomedicine **3**, 797–812 (2008).
[16] E. Sassaroli and K. Hynynen, "Forced linear oscillations of microbubbles in blood capillaries," J. Acoust. Soc. Am. **115**, 3235–3243 (2004).
[17] C. K. Turangan, A. R. Jamaluddin, G. J. Ball, and T. G. Leighton, "Free-Lagrange simulations of the expansion and jetting collapse of air bubbles in water," J. Fluid Mech. **598**, 1–25 (2008).
[18] H. Miao and S. M. Gracewski, "Coupled FEM and BEM code for simulating acoustically excited bubbles near deformable structures," Comput. Mech. **42**, 95–106 (2008).
[19] A. Prosperetti and G. Tryggvason, *Computational Methods for Multiphase Flow* (Cambridge University Press, Cambridge, 2008).
[20] S. Temkin, *Suspension Acoustics* (Cambridge University Press, New York, 2005).
[21] D. A. Drew, "Mathematical model of two-phase flow," Annu. Rev. Fluid Mech. **15**, 261–291 (1983).
[22] R. Jackson, "Locally averaged equations of motion for a mixture of identical spherical particles and a Newtonian fluid," Chem. Eng. Sci. **52**, 2457–2469 (1997).
[23] D. D. Joseph and T. S. Lundgren, "Ensemble averaged and mixture theory equations for incompressible fluid-particle suspensions," Int. J. Multiphase Flow **16**, 35–42 (1990).
[24] R. I. Nigmatulin, "Spatial averaging in the mechanics of heterogeneous and dispersed systems," Int. J. Multiphase Flow **5**, 353–385 (1979).
[25] D. Z. Zhang and A. Prosperetti, "Averaged equations for inviscid dispersed two-Phase flow," J. Fluid Mech. **267**, 185–219 (1994).
[26] D. Z. Zhang and A. Prosperetti, "Momentum and energy equations for disperse two-phase flows and their closure for diluted suspensions," Int. J. Multiphase Flow **23**, 425–453 (1997).
[27] H. C. van de Hulst, *Light Scattering by Small Particles* (Dover, New York, 1981).
[28] A. Prosperetti, L. A. Crum, and K. W. Commander, "Nonlinear bubble dynamics," J. Acoust. Soc. Am. **83**, 502–514 (1988).
[29] A. Prosperetti and Y. Hao, "Modeling of spherical gas bubble oscillations and sonoluminescence," Philos. Trans. R. Soc. London, Ser. A **357**, 203–223 (1999).
[30] M. S. Plesset and A. Prosperetti, "Bubble dynamics and cavitation," Annu. Rev. Fluid Mech. **9**, 145–185 (1977).
[31] E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).
[32] P. B. Johnson and R. W. Christy, "Optical constants of the noble metal," Phys. Rev. B **6**, 4370–4379 (1972).
[33] C. Sonnichsen, T. Franzl, T. Wilk, G. von Plessen, and J. Feldmann, "Plasmon resonances in large noble-metal clusters," New J. Phys. **4**, 93.1–93.8 (2002).
[34] *American Institute of Physics Handbook* (McGraw-Hill, New York, 1972).
[35] N. B. Vargaftik, *Handbook of Thermophysical Properties of Gases and Liquids* (Nauka, Moscow, 1972).
[36] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon, Oxford, 1987).
[37] S. Hilgenfeldt and D. Lohse, "The acoustics of diagnostic microbubbles: Dissipative effects and heat deposition," Ultrasonics **38**, 99–104 (2000).
[38] L. D. Rozenberg, "Absorption in the medium," in *Sources of High-Intensity Ultrasound Vol. 1*, edited by L. Z. Rozenberg (Plenum, New York, 1969), pp. 263–287.

Erratum: Detection of time-varying harmonic amplitude alterations due to spectral interpolations between musical instrument tones [J. Acoust. Soc. Am. 125, 492 (2009)]

Andrew B. Horner
Department of Computer Science, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

James W. Beauchamp
School and Music and Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

Richard H. Y. So
Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

(Received 20 August 2009; accepted 25 August 2009)
[DOI: 10.1121/1.3238163]

PACS number(s): 43.75.Cd, 43.66.Jh, 43.10.Vx

In the abstract, the sentence beginning "Among primary instruments…" should be replaced by "Among primary instruments, it was found that changes to horn and bassoon were most easily discriminable, while changes to saxophone and trumpet timbres were least discriminable."

On p. 496, Eq. (8) should be replaced by

$$\varepsilon = \frac{1}{N}\sum_{n=0}^{N-1}\sqrt{\frac{\sum_{k=1}^{K}(A_k(t_n)-A'_k(t_n))^2}{\sum_{k=1}^{K}A_k^2(t_n)}}. \tag{8}$$

The authors regret these errors.

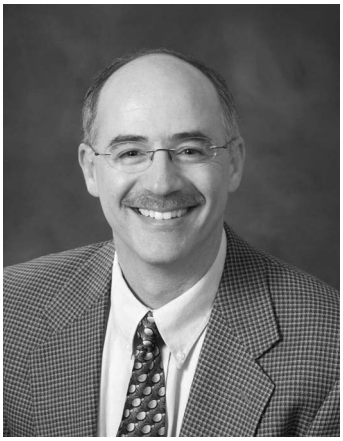
ACOUSTICAL NEWS

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

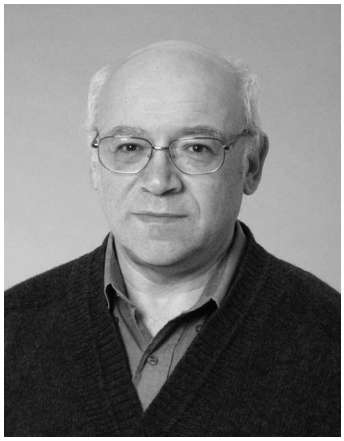
Editor's Note: *Readers of this journal are encouraged to submit news items on awards, appointments, and other activities about themselves or their colleagues. Deadline dates for news and notices are 2 months prior to publication.*

New Fellows of the Acoustical Society of America

Constantin C.-Coussios—For contributions to therapeutic ultrasound



Michael F. Insana—For contributions to biomedical ultrasound



Boris Katsnelson—For contributions to shallow water acoustics

Philip C. Loizou—For contributions to cochlear implant signal processing

Dennis A. Paoletti—For contributions to acoustic design and outreach

Shin-ichiro Umemura—For contributions to diagnostic and therapeutic ultrasound devices

Report of the Technical Committee on Animal Bioacoustics

(See September issue for reports of other Technical Committees)

There is much to report on the activities of the Technical Committee on Animal Bioacoustics (TCAB) and its members during the period from July 2008 and May 2009.

The 156th meeting was held 14–18 November 2008 at the Doral Golf Resort and Spa in Miami, Florida. TCAB sponsored and cosponsored numerous sessions as follows: (1, 2, 3) Marine Mammal Acoustics in Honor of Sam Ridgway I, II, and III organized by Whitlow Au and Dorian Houser drew a total of 35 papers, 16 invited and 18 contributed. Session chairs were Whitlow Au (Hawaii Inst. of Marine Biology), James Finneran (U.S. Navy Marine Mammal Program, Space and Naval Warfare Systems Ctr., San Diego), and Dorian S. Houser (Biomimetica). (4) Acoustics of Manatees and Alligators and Other Topics was organized and chaired by Jennifer L. Miksis-Olds (Pennsylvania State Univ.) and Ann E. Bowles (Hubbs Sea World Research Inst.) and included 11 papers, 5 invited and 6 contributed. (5) Animal Bioacoustics General Session chaired by David C. Swanson (Penn State Univ.) included 10 contributed papers on topics such as bats,

 0001-4966/2009/126(5)/2815/3/$25.00

zebra finches, budgerigars, insects, and an online archive for natural sounds. The TCAB also cosponsored Advances in Measurement and Noise and Noise Effects on Humans and Non-Human Animals in the Environment II with the Technical Committee on Noise. The session, which included 7 invited and 1 contributed paper was organized by Ann E. Bowles, Brigitte Schulte-Fortkamp (Technical Univ. of Berlin), and Kurt Fristrip. Dorian S. Houser served as the Animal Bioacoustics representative to the Technical Program Organizing Meeting. Hsiao-Wei Tu, University of Maryland, was named recipient of the Best Student Paper Award in Animal Bioacoustics for his paper "Categorization of budgerigar (*Melopsittacus undulatus*) warble elements."

The 157th meeting was held 16–20 May 2009 at the Hilton Portland and Executive Tower in Portland, Oregon. TCAB sponsored and cosponsored 14 sessions. (1 & 2) Fish Bioacoustics: Sensory Biology, Sound Production, and Behavior of Acoustic Communication in Fishes I and II, cosponsored by Psychological and Physiological Acoustics and organized by Joseph A. Sisneros, was chaired by Joseph A. Sisneros (Univ. of Washington), Richard R. Fay (Loyola Univ.), and David Zeddies (Marine Acoustics Inc.), and included 23 invited and 2 contributed papers. (3 & 4) An Integration of Bioacoustics, Neuronal Responses, and Behavior I and II (organized and chaired by Terry Takahashi (Univ. of Oregon) drew a total of 8 invited and 5 contributed papers and concluded with a panel discussion. (5 & 6) General Topics in Animal Bioacoustics I and II, chaired by Holger Klinck and David K. Mellinger respectively (both from Oregon State Univ.) included 20 papers total. (7–10) Autonomous Remote Monitoring Systems for Marine Animals I, II, III, and IV (cosponsored by Acoustical Oceanography and organized by Marc Lammers) was chaired by Marc Lammers, (Hawaii Inst. of Marine Biology—Sessions I and III), Kathleen C. Stafford (Univ. of Washington—Session II), and Catherine L. Berchock, (Alaska Fisheries Science Ctr./NOAA—Session IV) and drew a total of 40 papers including 14 invited. (11) Signal Processing Techniques for Subtle or Complex Acoustic Features of Animal Calls (cosponsored by Signal Processing in Acoustics and organized by Ann Bowles and Sean K. Lehman) was chaired by Ann E. Bowles and drew 13 papers, 11 of which were invited. TCAB also cosponsored the following sessions. (12) Bioacoustic Metrics and the Impact of Noise on the Natural Environment (Noise the primary sponsor with ASACOS) was chaired by Michael Stocker (Ocean Conservation Research) with 12 papers, 3 of which were invited. (13) Acoustic Backscattering from Marine Life in the Ocean (Acoustical Oceanography the primary sponsor) was chaired by Christopher D. Jones (Univ. of Washington) with 11 contributed papers. (14) Source/Filter Interaction in Biological Sound Production, with Speech Communication and Musical Acoustics) chaired by Ingo R. Titze with 13 papers, 5 of which were invited. Holger Klinck was representative to the Technical Program Organizing Meeting. Best Student Paper Awards were presented to: First prize: Mary Bates, Brown University, for the paper "Is rejection of clutter achieved by disrupting perception of delay in bat sonar?" and Second Prize: Asila Ghoul, University of California, Santa Cruz for the paper "Auditory temporal summation in pinnipeds."

Animal Bioacoustics topics were very well represented in the journal publications of the Society. There were over 50 papers published in the *Journal of the Acoustical Society of America* and 7 in *JASA Express Letters*—the ASA's open access, rapid-publication journal—during the period covered by this report.

Associate Editors serving in the area of Animal Bioacoustics are Whitlow W. L. Au, Mardi C. Hastings, Michael J. Owren, James M. Simmons for *JASA*; Cynthia Moss for *JASA Express Letters*, and Richard R. Fay for *Proceedings of Meetings on Acoustics*. TCAB extends its thanks and congratulations to Floyd Dunn who announced his retirement as Associate Editor of JASA. Floyd handled the review process for many papers in Animal Bioacoustics during his 33-years of service as Associate Editor.

In 2008, the ASA added a new title to its collection of e-books, *Hearing in Vertebrates: A Psychophysics Databook* by Richard R. Fay. This book presents psychophysical data on vertebrate hearing obtained from the published literature. In addition to data on hearing sensitivity, discrimination, and directional hearing, data are included on hearing development and infant hearing, aspects of echolocation, and the psychophysics of electrical stimulation of the auditory system. It is available for purchase and immediate download in pdf format at the ASA Store http://asastore.aip.org/.

TCAB is represented by its members on various ASA committees that help in conducting the business of the Society. They include the Associate Editors mentioned above who serve on the Editorial Board, James M. Simmons on the Medals and Awards Committee, Richard R. Fay who replaced Andrea Simmons in 2008 on the Membership Committee, Alison Stimpert and Mary E. Bates on the Student Council, and Anne E. Bowles on the ASA Committee on Standards. Although technical committees are not individually represented on the Public Relations Committee, over 15 authors of Miami and Portland meeting papers on animal bioacoustics topics prepared lay-language versions of their papers to be used for outreach by ASA to science writers and the general public. These can be viewed at the ASA World Wide Press Room at http://www.acoustics.org/press/index.html.

The Animal Bioacoustics webpage has been and is available at http://www.animalbioacoustics.org/. Thanks to David Mellinger and Holger Klinck for the redesign and maintenance of the site where visitors can find information about animal bioacoustics topics such as new books, conferences, and hardware and software used in recording.

TCAB members also participate in the Standards activities of the Society through its representative on ASACOS and through a new Standards Committee: Accredited Standards Committee S3, Subcommittee 1, Animal Bioacoustics. David K. Delaney of USA CERL serves as Chair and Mardi C. Hastings, serves as Vice Chair. This committee's scope covers standards, specifications, methods of measurement and test, instrumentation and terminology in the field of psychological and physiological acoustics, including aspects of general acoustics, which pertain to biological safety, tolerance and comfort of non-human animals, including both risk to individual animals and to the long-term viability of populations. Animals to be covered may potentially include commercially-grown food animals; animals harvested for food in the wild; pets; laboratory animals; exotic species in zoos, oceanaria or aquariums; or free-ranging wild animals. There are currently two active standards writing groups functioning under the auspices of the Committee including S3/SC 1/WG1 Animal Bioacoustics Terminology and S3/SC 1/WG2 Effects of Sound on Fish and Turtles. For additional details about Standards activities of this group, please visit http://www.acosoc.org/standards/S3-SC1/S3-SC1.htm.

Several members of the ASA animal bioacoustics community have been recognized for their accomplishments during the past year. They include Kelly J. Benoit-Bird who was named recipient of the 2009 R. Bruce Lindsay Award "For contributions in marine ecological acoustics;" Patrick Moore and Aaron Thode who were elected Fellows of the Society; Alison Stimpert who was named recipient of the 2009-10 Frederick V. Hunt Postdoctoral Research Fellow in Acoustics; and Kathleen Vigness-Raposa, Gail Scowcroft, Christopher Knowlton, and Peter Worcester who received ASA's 2008 Science Writing Award for Media Other than Articles for the "Discovery of Sound in the Sea" website. ASA's 2009 Robert W. Young Grant for Undergraduate Research was awarded for a project involving the study of animals, namely, Cody Brack, an undergraduate student at St. Mary's College of Maryland, for his project titled "A Macro-Level Assessment of Zebrafish." Congratulations are also extended to Whitlow W.L. Au, former chair of the TCAB, who was elected President-Elect and who assumed the office of President in May 2009 for a one-year term.

Further evidence of the TCAB's outreach and influence on the animal bioacoustics committee, both in the U.S. and abroad, is the sponsorship and cosponsorship of symposia and other meetings. The ASA, along with Oregon State University (OSU) organized and presented the Second International Conference on Acoustic Communication by Animals which was held August 12–15, 2008 in Corvallis, Oregon. The organizers of the symposium were David K. Mellinger and Sarah Heimlich; this conference followed the successful First International Conference on Acoustic Communication by Animals, held at the University of Maryland in July 2003. Over 200 people attended the symposium which was cosponsored by Office of Naval Research, International Commission for Acoustics, National Park Service, National Oceanographic and Atmospheric Administration, Marine Mammal Commission, Engineer Research and Development Center, Dept. of the Interior Minerals Management Service.

The Keynote Speakers were Peter Marler who spoke on acoustic communication and learning and Peter Slater whose presentation was on a Tropical Perspective on Bird Song. Invited Speakers were Whitlow Au, Andrew Bass, Eliot Brenowitz, Robert Dooling, Gunter Ehret, Richard Fay, Albert Feng, Tecumseh Fitch, Kurt Fristrup, Peter Narins, Kazuo Okanoya, Arthur Popper, Denise Risch, Caitlin O'Connell-Rodwell, Ron Schusterman, Andrea Megela Simmons, James Simmons, Joseph Sisneros, AnnMarie Surlykke, Terry Takahashi, Peter Tyack, and Sophie Van Parijs who covered the vast array of animal communication topics including marine mammals, birds, fish, land mammals, frogs, pinnipeds, and bats. ASA also cosponsored the 5th Animal Sonar Symposium held in Kyoto, Japan, September 14–18, 2009. Several ASA members served on the organizing committee for this symposium including Whitlow W. L. Au, Lee A. Miller, Cynthia F. Moss, Paul E. Nachtigall, Hiroshi Riquimaroux, James A. Simmons, Jeanette A.

Thomas, and Tomonari Akamatsu. Details of the conference were not available at press time and will be reported in a future TCAB report.

The Committee expresses its heartfelt thanks to all the volunteer members mentioned above who have participated in ASA activities on behalf of the Committee.

David K. Mellinger of Oregon State University, was elected Chair of the Technical Committee on Animal Bioacoustics for a three-year term to Spring 2012. We hope that many readers of this report will join him and other colleagues in the ASA animal bioacoustics community at upcoming meetings of the ASA.

RICHARD R. FAY
Chair 2006–2009

Calendar of Meetings and Congresses

2009

5-6 November	Dübendorf, Switzerland. Swiss Acoustical Society Autumn Meeting. Web: www.sga-ssa.ch
11-13 November	Miyagi, Japan. 1st International Workshop on Principles and Applications of Spatial Hearing. Web: www.riec.tohoku.ac.jp/IWPASH/
18-20 November	Kyoto, Japan. 30th Symposium on Ultrasonics Electronics. Web: www.use-jp.org/USE2009/en/index.html
18-20 November	Kochi, India. International Symposium on Ocean Electronics Sympol 2009. Web: http://sympol.cusat.ac.in/
23-25 November	Adelaide, Australia. Australian Acoustics Society National Conference. Web: www.acoustics.asn.au/joomla/acoustics-2009.html

2010

6-9 January	Sanya, China. 2nd International Conference on Vibro-Impact Systems. Web: www.neu.edu.cn
8-11 March	Berlin, Germany. Meeting of the German Association for Acoustics DAGA 2010. Web: www.daga-tagung.de/2010
15-19 March	Dallas, TX, USA. International Conference on Acoustics, Speech, and Signal Processing. Web: http://icassp2010.org
7–9 April	Cambridge, UK. David Weston Sonar Performance Assessment Symposium. E-mail: michael.ainslie@tno.nl
19-23 April	Baltimore, MD, USA. Joint meeting: 159th Meeting of the Acoustical Society of America and Noise Con 2010. Web: http://asa.aip.org/meetings.html
27-30 April	Ghent, Belgium. Institute of Acoustics/Belgian Acoustical Association Joint Meeting. Web: www.ioa.org.uk/viewupcoming.asp
9-11 June	Aalborg, Denmark. 14th Conference on Low Frequency Noise and Vibration. Web: http://lowfrequency2010.org
13-16 June	Lisbon, Portugal. INTERNOISE2010. Web: www.internoise2010.org
5-9 July	Istanbul, Turkey. 10th European Conference on Underwater Acoustics. Web: http://ecua-2010-istanbul.org
23-27 August	Seattle, USA. 11th International Conference on Music Perception and Cognition. Web: http://www.musicperception.org/resources/ICMPC11_Flyer.pdf
23-27 August	Sydney, Australia. International Congress on Acoustics 2010. Web: www.ica2010sydney.org
29-31 August	Melbourne, Australia. International Symposium on Room Acoustics (ISRA2010). Web: http://web.arch.usyd.edu.au/~densil/ISRA/
14-18 September	Kyoto, Japan. 5th Animal Sonar Symposium. Web: http://cse.fra.affrc.go.jp/akamatsu/AnimalSonar.html
15-18 September	Ljubljania, Slovenia. Alp-Adria-Acoustics Meeting joint with EAA. E-mail: mirko.cudina@fs.uni-lj.si
26-30 September	Makuhari, Japan. Interspeech 2010—ICSLP. Web: www.interspeech2010.org
14-16 October	Niagara-on-the Lake, Ont., Canada. Acoustics Week in Canada. Web: http://caa-aca.ca/E/index.html
11-14 October	San Diego, California, USA. IEEE 2010 Ultrasonics Symposium. E-mail: bpotter@vectron.com
15-19 November	Cancun, Mexico. Second Pan-American/Iberian Meeting on Acoustics (Joint meeting of the Acoustical Society of America, Iberoamerican Congress of Acoustics, Mexican Congress on Acoustics. Web: http://asa.aip.org/meetings.html
19-20 November	Brighton, UK. Reproduced Sound 25. Web: www.ica.org.uk/viewupcoming.asp

2011

22-27 May	Prague, Czech Republic. International Conference on Acoustics, Speech, and Signal Processing (IEEE ICASSP 2011). Web: http://www.icassp2011.com/
27 June-1 July	Aalborg, Denmark. Forum Acusticum 2011. Web: www.fa2011.org
16-21 July	Williamstown, Massachusetts, USA. 11th International Mechanics of Hearing Workshop. Web: www.mechanicsofhearing.org/
24-28 July	Tokyo, Japan. 19th International Symposium on Nonlinear Acoustics (ISNA 19). Web: TBA
27-31 August	Florence, Italy. Interspeech 2011. Web: www.interspeech2011.org
04-07 September	Osaka, Japan. Internoise 2011. Web: www.internoise2011.com
5-8 September	Gdansk, Poland. International Congress on Ultrasonics. Web: http://icu2011.ug.edu.pl/index.html

2013

2-7 June	Montréal, Canada. 21st International Congress on Acoustics (ICA 2013) (Joint meeting: International Congress on Acoustics, Acoustical Society of America, Canadian Acoustical Association). Web: www.ica2013montreal.org

ACOUSTICAL STANDARDS NEWS

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

Paul D. Schomer, Standards Director
Schomer and Associates, 2117 Robert Drive, Champaign, IL 61821 [Tel.: (217) 359-6602; Fax: (217) 359-3303; e-mail: Schomer@SchomerAndAssociates.com]

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

Comments are welcomed on all material in Acoustical Standards News.

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

Standards Meetings Calendar—National

Accredited Standards Committees S1, Acoustics; S2, Mechanical Vibration and Shock; S3, Bioacoustics; S3/SC 1, Animal Bioacoustics; and S12, Noise, along with the U.S. Technical Advisory Groups to ISO/TC 43, ISO/TC 43/SC 1, ISO/TC 108 and its five Subcommittees, and IEC/TC 29, ASACOS and the Standards Plenary Group will meet in conjunction with the Joint 159th ASA Meeting and Noise-Con 2010 to be held in Baltimore, Maryland, **19–23 April 2010**. Additional details will be provided when available.

ASTM E-33 "Committee on Building and Environmental Acoustics" has scheduled two meetings as follows:

- **17–18 May 2010** in St. Louis, MO; and
- **11–12 October 2010** in San Antonio, TX.

For more information, visit www.astm.org.

Standards Meetings Calendar—International

ISO/TC 108/SC5 Condition Monitoring and Diagnostics of Machines will meet in Paris, France the week of **20 September 2010**.

Recent International Meetings

ISO/TC 108/SC 5 met near Copenhagen in June 2009. The U.S. was represented by a delegation of nine experts in the field of "Condition monitoring and diagnostics of machines," led by ASA member Dr. David J. Vendittis of Florida Atlantic University. Overall, the meeting attracted more than 40 experts from 12 countries. Ten of SC 5's eleven working groups met during the week along with two advisory groups. SC 5's work programme covers a range of techniques applied in CM&D programs. A new working group on Condition Monitoring of Wind Turbines was formed.

ISO/TC 108/SC 4, "Human exposure to mechanical vibration and shock," met in Las Vegas in September. This was the first time the U.S. has hosted this subcommittee since the 1980's. The meeting was funded by sponsorship from the following members of the U.S. TAG: U.S. Department of Transportation, NIOSH, Commercial Vehicle Group, The Vibration Institute, Caterpillar, NSWC Panama City, and Occupational Medicine Consultants. More than 40 international experts on the topic assembled to discuss issues including "hand transmitted vibration," "biodynamic modeling," "vibrotactile perception," and "human exposure to repetitive mechanical shock" among others.

Anyone interested in learning more about ISO/TC 108/SC 4 or SC 5 is invited to contact the Standards Manager. Information is also available on the ASA website by clicking the "Standards Info" button.

Call for Data on Biodynamic Response to Whole-body Vibration

ISO/TC 108/SC 4/WG 5, Biodynamic modeling, is seeking experimental data to extend the applicability of ISO 5982 (the standard describing biodynamic response of seated persons) to more practical situations, e.g., car, truck, and train seats. To submit data, please contact Dr. A.J. Brammer, e-mail: tony.brammer@nrc-cnrc.gc.ca.

A listing of the Accredited Standards Committees, their Working Groups, their Chairs (listed in parenthesis), and published standards; Chair and Vice Chair of the ASA Committees on Standards (ASACOS); and the U.S. Technical Advisory Group (TAG) Chairs for the International Standards Committees are given here for reference:

Accredited Standards Committee on Acoustics, S1

(P. Battenberg, Chair; R.J. Peppin, Vice Chair)

Scope: Standards, specifications, methods of measurement and test, and terminology in the field of physical acoustics including architectural acoustics, electroacoustics, sonics and ultrasonics, and underwater sound, but excluding those aspects which pertain to biological safety, tolerances, and comfort.

S1 Working Groups

S1/Advisory—Advisory Planning Committee to S1 (P. Battenberg, Chair; R.J. Peppin, Vice Chair);

S1/WG1—Standard Microphones and their Calibration (V. Nedzelnitsky);

S1/WG2—Attenuation of Sound in the Atmosphere (A.H. Marsh);

S1/WG4—Measurement of Sound Pressure Levels in Air (VACANT, Chair; E. Dunens, Vice Chair);

S1/WG5—Band Filter Sets (A.H. Marsh);

S1/WG9—Calibration of Underwater Electroacoustic Transducers (R.M. Drake);

S1/WG16—FFT Acoustical Analyzers (R.L. McKinley);

S1/WG17—Sound Level Meters and Integrating Sound Level Meters (G.R. Stephany);

S1/WG19—Insertion Loss of Windscreens (A.J. Campanella);

S1/WG20—Ground Impedance (Measurement of Ground Impedance and Attenuation of Sound Due to the Ground) (K. Attenborough, Chair; J. Sabatier, Vice Chair);

S1/WG22—Bubble Detection and Cavitation Monitoring (Vacant);

S1/WG27—Acoustical Terminology (J.S. Vipperman).

S1 Standards on Acoustics

ANSI S1.1-1994 (R 2004) American National Standard Acoustical Terminology.

ANSI S1.4-1983 (R 2006) American National Standard Specification for Sound Level Meters. This Standard includes **ANSI S1.4A-1985 (R 2006)** Amendment to ANSI S1.4-1983.

ANSI S1.6-1984 (R 2006) American National Standard Preferred Frequencies, Frequency Levels, and Band Numbers for Acoustical Measurements.

ANSI S1.8-1989 (R 2006) American National Standard Reference Quantities for Acoustical Levels.

ANSI S1.9-1996 (R 2006) American National Standard Instruments for the Measurement of Sound Intensity.

ANSI/ASA S1.11-2004 (R 2009) American National Standard Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters.

ANSI S1.13-2005 American National Standard Measurement of Sound Pressure Levels in Air.

ANSI/ASA S1.14-1998 (R 2008) American National Standard Recommendations for Specifying and Testing the Susceptibility of Acoustical Instruments to Radiated Radio-Frequency Electromagnetic Fields, 25 MHz to 1 GHz.

ANSI S1.15/Part 1-1997 (R 2006) American National Standard Measurement Microphones, Part 1: Specifications for Laboratory Standard Microphones.

ANSI S1.15/Part 2-2005 American National Standard Measurement Microphones, Part 2: Primary Method for Pressure Calibration of Laboratory Standard Microphones by the Reciprocity Technique.

ANSI S1.16-2000 (R 2005) American National Standard Method for Measuring the Performance of Noise Discriminating and Noise Canceling Microphones.

ANSI S1.17/Part 1-2004 American National Standard Microphone Windscreens—Part 1: Measurements and Specification of Insertion Loss in Still or Slightly Moving Air.

ANSI S1.18-1999 (R 2004) American National Standard Template Method for Ground Impedance.

ANSI S1.20-1988 (R 2003) American National Standard Procedures for Calibration of Underwater Electroacoustic Transducers.

ANSI S1.22-1992 (R 2007) American National Standard Scales and Sizes for Frequency Characteristics and Polar Diagrams in Acoustics.

ANSI S1.24 TR-2002 (R 2007) ANSI Technical Report Bubble Detection and Cavitation Monitoring.

ANSI S1.25-1991 (R 2007) American National Standard Specification for Personal Noise Dosimeters.

ANSI/ASA S1.26-1995 (R 2009) American National Standard Method for Calculation of the Absorption of Sound by the Atmosphere.

ANSI S1.40-2006 American National Standard Specifications and Verification Procedures for Sound Calibrators.

ANSI S1.42-2001 (R 2006) American National Standard Design Response of Weighting Networks for Acoustical Measurements.

ANSI S1.43-1997 (R 2007) American National Standard Specifications for Integrating-Averaging Sound Level Meters.

Accredited Standards Committee on Mechanical Vibration and Shock, S2

(A.T. Herfat, Chair; C.F. Gaumond, Vice Chair)

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

S2 Working Groups

S2/WG1—S2 Advisory Planning Committee (A.T. Herfat, Chair; C.F. Gaumond, Vice Chair);

S2/WG2—Terminology and Nomenclature in the Field of Mechanical Vibration and Shock and Condition Monitoring and Diagnostics of Machines (D.J. Evans);

S2/WG3—Signal Processing Methods (T.S. Edwards);

S2/WG4—Characterization of the Dynamic Mechanical Properties of Viscoelastic Polymers (W.M. Madigosky, Chair; J. Niemiec, Vice Chair);

S2/WG5—Use and Calibration of Vibration and Shock Measuring Instruments (D.J. Evans, Chair; B.E. Douglas, Vice Chair);

S2/WG6—Vibration and Shock Actuators (G. Booth);

S2/WG7—Acquisition of Mechanical Vibration and Shock Measurement Data (B.E. Douglas);

S2/WG8—Analysis Methods of Structural Dynamics (M. Mezache);

S2/WG9—Training and Accreditation (R. Eshleman, Chair; D. Corelli, Vice Chair);

S2/WG10—Measurement and Evaluation of Machinery for Acceptance and Condition (R.L. Eshleman, Chair; H. Pusey, Vice Chair);

S2/WG10/Panel 1—Balancing (R.L. Eshleman);

S2/WG10/Panel 2—Operational Monitoring and Condition Evaluation (R. Bankert);

S2/WG10/Panel 3—Machinery Testing (R.L. Eshleman);

S2/WG10/Panel 4—Prognosis (A.J. Hess);

S2/WG10/Panel 5—Data Processing, Communication, and Presentation (K. Bever);

S2/WG11—Measurement and Evaluation of Mechanical Vibration of Vehicles (Vacant);

S2/WG12—Measurement and Evaluation of Structures and Structural Systems for Assessment and Condition Monitoring (M. Mezache);

S2/WG13—Shock Test Requirements for Shelf-Mounted and Other Commercial Electronics Systems (B. Lang);

S2/WG39 (S3)—Human Exposure to Mechanical Vibration and Shock (D.D. Reynolds, Chair; R. Dong, Vice Chair).

S2 Standards on Mechanical Vibration and Shock

ANSI/ASA S2.1-2009/ISO 2041:2009 American National Standard Vibration and Shock—Vocabulary (Nationally Adopted International Standard).

ANSI S2.2-1959 (R 2006) American National Standard Methods for the Calibration of Shock and Vibration Pickups.

ANSI S2.4-1976 (R 2004) American National Standard Method for Specifying the Characteristics of Auxiliary Analog Equipment for Shock and Vibration Measurements.

ANSI S2.8-2007 American National Standard Technical Information Used for Resilient Mounting Applications.

ANSI/ASA S2.9-2008 American National Standard Parameters for Specifying Damping Properties of Materials and System Damping.

ANSI S2.16-1997 (R 2006) American National Standard Vibratory Noise Measurements and Acceptance Criteria of Shipboard Equipment.

ANSI S2.19-1999 (R 2004) American National Standard Mechanical Vibration—Balance Quality Requirements of Rigid Rotors, Part 1: Determination of Permissible Residual Unbalance, Including Marine Applications.

ANSI S2.20-1983 (R 2006) American National Standard Estimating Air Blast Characteristics for Single Point Explosions in Air, with a Guide to Evaluation of Atmospheric Propagation and Effects.

ANSI S2.21-1998 (R 2007) American National Standard Method for Preparation of a Standard Material for Dynamic Mechanical Measurements.

ANSI S2.22-1998 (R 2007) American National Standard Resonance Method for Measuring the Dynamic Mechanical Properties of Viscoelastic Materials.

ANSI S2.23-1998 (R 2007) American National Standard Single Cantilever Beam Method for Measuring the Dynamic Mechanical Properties of Viscoelastic Materials.

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

ANSI/ASA S2.25-2004 (R 2009) American National Standard Guide for the Measurement, Reporting, and Evaluation of Hull and Superstructure Vibration in Ships.

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

ANSI S2.27-2002 (R 2007) American National Standard Guidelines for the Measurement and Evaluation of Vibration of Ship Propulsion Machinery.

ANSI/ASA S2.28-2009 American National Standard Guide for the Measurement and Evaluation of Broadband Vibration of Surface Ship Auxiliary Rotating Machinery.

ANSI/ASA S2.29-2003 (R 2008) American National Standard Guide for the Measurement and Evaluation of Vibration of Machine Shafts on Shipboard Machinery.

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

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

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

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

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

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

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

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

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

ANSI/ASA S2.62-2009 American National Standard Shock Test Requirements for Equipment in a Rugged Shock Environment.

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

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

ANSI S2.72/Part 1-2002 (R 2007)/ISO 2631-1:1997 *(Redesignation of ANSI S3.18/Part 1-2002/ISO 2631-1:1997)* American National Standard Mechanical vibration and shock-Evaluation of human exposure to whole-body vibration—Part 1: General requirements (Nationally Adopted International Standard).

ANSI S2.72/Part 4-2003 (R 2007)/ISO 2631-4:2001 *(Redesignation of ANSI S3.18/Part 4 -2003 / ISO 2631-4:2001)* American National Standard Mechanical vibration and shock—Evaluation of human exposure to whole-body vibration—Part 4: Guidelines for the evaluation of the effects of vibration and rotational motion on passenger and crew comfort in fixed-guideway transport systems (Nationally Adopted International Standard).

ANSI S2.73-2002 (R 2007) / ISO 10819:1996 *(Redesignation of ANSI S3.40-2002 / ISO 10819:1996)* American National Standard Mechanical vibration and shock—Hand-arm vibration—Method for the measurement and evaluation of the vibration transmissibility of gloves at the palm of the hand (Nationally Adopted International Standard).

Accredited Standards Committee on Bioacoustics, S3

(C.A. Champlin, Chair; D.A. Preves, Vice Chair)

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

S3 Working Groups

S3/Advisory—Advisory Planning Committee to S3 (C.A. Champlin, Chair; D.A. Preves, Vice Chair);

S3/WG35—Audiometric Equipment (R.L. Grason);

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

S3/WG37—Coupler Calibration of Earphones (C.J. Struck);

S3/WG39—Human Exposure to Mechanical Vibration and Shock—Parallel to ISO/TC 108/SC 4 (D.D. Reynolds, Chair; R. Dong, Vice Chair);

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

S3/WG48—Hearing Aids (D.A. Preves);

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

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

S3/WG62—Impulse Noise with Respect to Hearing Hazard (G.R. Price);

S3/WG67—Manikins (M.D. Burkhard);

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

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

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

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

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

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

S3/WG84—Otoacoustic Emissions (Vacant);

S3/WG88—Standard Audible Emergency Evacuation and Other Signals (R. Boyer);

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

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

S3 Liaison Group

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

S3 Standards on Bioacoustics

ANSI/ASA S3.1-1999 (R 2008) American National Standard Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms.

ANSI/ASA S3.2-2009 American National Standard Method for Measuring the Intelligibility of Speech over Communication Systems.

ANSI S3.4-2007 American National Standard Procedure for the Computation of Loudness of Steady Sounds.

ANSI S3.5-1997 (R 2007) American National Standard Methods for Calculation of the Speech Intelligibility Index.

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

ANSI/ASA S3.7-1995 (R 2008) American National Standard Method for Coupler Calibration of Earphones.

ANSI S3.13-1987 (R 2007) American National Standard Mechanical Coupler for Measurement of Bone Vibrators.

ANSI/ASA S3.20-1995 (R 2008) American National Standard Bioacoustical Terminology.

ANSI/ASA S3.21-2004 (R 2009) American National Standard Methods for Manual Pure-Tone Threshold Audiometry.

ANSI/ASA S3.22-2009 American National Standard Specification of Hearing Aid Characteristics.

ANSI S3.25-2009 American National Standard for an Occluded Ear Simulator.

ANSI S3.35-2004 American National Standard Method of Measurement of Performance Characteristics of Hearing Aids under Simulated Real-Ear Working Conditions.

ANSI S3.36-1985 (R 2006) American National Standard Specification for a Manikin for Simulated *in situ* Airborne Acoustic Measurements.

ANSI S3.37-1987 (R 2007) American National Standard Preferred Earhook Nozzle Thread for Postauricular Hearing Aids.

ANSI S3.39-1987 (R 2007) American National Standard Specifications for Instruments to Measure Aural Acoustic Impedance and Admittance (Aural Acoustic Immittance).

ANSI S3.41-1990 (R 2008) American National Standard Audible Emergency Evacuation Signal.

ANSI S3.42-1992 (R 2007) American National Standard Testing Hearing Aids with a Broad-Band Noise Signal.

ANSI S3.44-1996 (R 2006) American National Standard Determination of Occupational Noise Exposure and Estimation of Noise-Induced Hearing Impairment.

ANSI/ASA S3.45-2009 American National Standard Procedures for Testing Basic Vestibular Function.

ANSI S3.46-1997 (R 2007) American National Standard Methods of Measurement of Real-Ear Performance Characteristics of Hearing Aids.

Accredited Standards Committee on Animal Bioacoustics, S3/SC1

(D.K. Delaney, Chair; M.C. Hastings, Vice Chair)

Scope: Standards, specifications, methods of measurement and test, instrumentation and terminology in the field of psychological and physiological acoustics, including aspects of general acoustics which pertain to biological safety, tolerance and comfort of non-human animals, including both risk to individual animals and to the long-term viability of populations.

Animals to be covered may potentially include commercially grown food animals; animals harvested for food in the wild; pets; laboratory animals; exotic species in zoos, oceanaria or aquariums; or free-ranging wild animals.

S3/SC1 Working Groups

S3/SC 1/WG1—Animal Bioacoustics Terminology (A.E. Bowles);

S3/SC 1/WG2—Effects of Sound on Fish and Turtles (R.R. Fay and A.N. Popper, Co-Chairs);

S3/SC 1/WG3—Underwater Passive Acoustic Monitoring for Bioacoustic Applications (A.M. Thode);

S3/SC 1/WG4—Description and Measurement of the Ambient Sound in Parks, Wilderness Areas, and Other Quiet and/or Pristine Areas (K. Fristrup and G.R. Stanley, Co-Chairs).

Accredited Standards Committee on Noise, S12

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

Scope: Standards, specifications, and terminology in the field of acoustical noise pertaining to methods of measurement, evaluation and control, including biological safety, tolerance and comfort, and physical acoustics as related to environmental and occupational noise.

S12 Working Groups

S12/Advisory—Advisory Planning Committee to S12 (W.J. Murphy and R.D. Hellweg);

S12/WG3—Measurement of Noise from Information Technology and Telecommunications Equipment (K. X. C. Man);

S12/WG11—Hearing Protector Attenuation and Performance (E.H. Berger);

S12/WG14—Measurement of the Noise Attenuation of Active and/or Passive Level Dependent Hearing Protective Devices (W.J. Murphy);

S12/WG15—Measurement and Evaluation of Outdoor Community Noise (P.D. Schomer);

S12/WG23—Determination of Sound Power (B.M. Brooks and J. Schmitt, Co-chairs);

S12/WG32—Revision of ANSI S12.7-1986 Methods for Measurement of Impulse Noise (W. Ahroon);

S12/WG38—Noise Labeling in Products (R.D. Hellweg);

S12/WG40—Measurement of the Noise Aboard Ships (S.P. Antonides, Chair; S.A. Fisher, Vice Chair);

S12/WG41—Model Community Noise Ordinances (L.S. Finegold, Chair; B.M. Brooks, Vice Chair);

S12/WG44—Speech Privacy (G.C. Tocci, Chair; D. Sykes, Vice Chair);

S12/WG45—Measurement of Occupational Noise Exposure from Telephone Equipment (K.A. Woo, Chair; L.A. Wilber, Vice Chair);

S12/WG46—Acoustical Performance Criteria for Relocatable Classrooms (T. Hardiman and P.D. Schomer, Co-Chairs);

S12/WG47—Underwater Noise Measurements of Ships (M. Bahtiarian, Chair; D.J. Vendittis, Vice Chair);

S12/WG48—Railroad Horn Sound Emission Testing (J. Erdreich, Chair; J.J. Earshen, Vice Chair);

S12/WG49—Noise from Hand-operated Power Tools, Excluding Pneumatic Tools (C. Hayden, Chair; B.M. Brooks, Vice Chair);

S12/WG50—Information Technology (IT) Equipment in Classrooms (R.D. Hellweg);

S12/WG51—Procedure for Measuring the Ambient Noise Level in a Room (J.G. Lilly);

S12/WG52—Revision of ANSI S12.60-2002 (S. Lind and P.D. Schomer, Co-Chairs).

S12 Liaison Groups

S12/L-1 IEEE 85 Committee for TAG Liaison—Noise Emitted by Rotating Electrical Machines (Parallel to ISO/TC 43/SC 1/WG 13) (R.G. Bartheld);

S12/L-2 Measurement of Noise from Pneumatic Compressors Tools and Machines (Parallel to ISO/TC 43/SC 1/WG 9) (Vacant);

S12/L-3 SAE Committee for TAG Liaison on Measurement and Evaluation of Motor Vehicle Noise (parallel to ISO/TC 43/SC 1/WG 8) (R.F. Schumacher);

S12/L-4 SAE Committee A-21 for TAG Liaison on Measurement and Evaluation of Aircraft Noise (J.D. Brooks);

S12/L-5 ASTM E-33 on Environmental Acoustics (to include activities of ASTM E33.06 on Building Acoustics, parallel to ISO/TC 43/SC 2 and ASTM E33.09 on Community Noise) (K.P. Roy);

S12/L-6 SAE Construction-Agricultural Sound Level Committee (I. Douell);

S12/L-7 SAE Specialized Vehicle and Equipment Sound Level Committee (T.M. Disch);

S12/L-8 ASTM PTC 36 Measurement of Industrial Sound (R.A. Putnam, Chair; B.M. Brooks, Vice Chair).

S12 Standards on Noise

ANSI S12.1-1983 (R 2006) American National Standard Guidelines for the Preparation of Standard Procedures to Determine the Noise Emission from Sources.

ANSI/ASA S12.2-2008 American National Standard Criteria for Evaluating Room Noise.

ANSI S12.3-1985 (R 2006) American National Standard Statistical Methods for Determining and Verifying Stated Noise Emission Values of Machinery and Equipment.

ANSI S12.5-2006/ISO 6926:1999 American National Standard Acoustics—Requirements for the Performance and Calibration of Reference Sound Sources Used for the Determination of Sound Power Levels (Nationally Adopted International Standard).

ANSI/ASA S12.6-2008 American National Standard Methods for Measuring the Real-Ear Attenuation of Hearing Protectors.

ANSI S12.7-1986 (R 2006) American National Standard Methods for Measurements of Impulse Noise.

ANSI/ASA S12.8-1998 (R 2008) American National Standard Methods for Determining the Insertion Loss of Outdoor Noise Barriers.

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

ANSI/ASA S12.9/Part 2-1992 (R 2008) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 2: Measurement of Long-Term, Wide-Area Sound.

ANSI/ASA S12.9/Part 3-1993 (R 2008) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 3: Short-Term Measurements with an Observer Present.

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

ANSI/ASA S12.9/Part 5-2007 American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound—Part 5: Sound Level Descriptors for Determination of Compatible Land Use.

ANSI/ASA S12.9/Part 6-2008 American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound—Part 6: Methods for Estimation of Awakenings Associated with Outdoor Noise Events Heard in Homes.

ANSI/ASA S12.10-2002 (R 2007)/ISO 7779:1999 American National Standard Acoustics—Measurement of airborne noise emitted by information technology and telecommunications equipment (Nationally Adopted International Standard).

ANSI/ASA S12.11/Part 1-2003 (R 2008)/ISO 10302:1996 (MOD) American National Standard Acoustics—Measurement of noise and vibration of small air-moving devices—Part 1: Airborne noise emission (Modified Nationally Adopted International Standard).

ANSI/ASA S12.11/Part 2—2003 (R 2008) American National Standard Acoustics—Measurement of Noise and Vibration of Small Air-Moving Devices—Part 2: Structure-Borne Vibration.

ANSI/ASA S12.12-1992 (R 2007) American National Standard Engineering Method for the Determination of Sound Power Levels of Noise Sources Using Sound Intensity.

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

ANSI/ASA S12.14-1992 (R 2007) American National Standard Methods for the Field Measurement of the Sound Output of Audible Public Warning Devices Installed at Fixed Locations Outdoors.

ANSI/ASA S12.15-1992 (R 2007) American National Standard For Acoustics—Portable Electric Power Tools, Stationary and Fixed Electric Power Tools, and Gardening Appliances—Measurement of Sound Emitted.

ANSI/ASA S12.16-1992 (R 2007) American National Standard Guidelines for the Specification of Noise of New Machinery.

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

ANSI/ASA S12.18-1994 (R 2009) American National Standard Procedures for Outdoor Measurement of Sound Pressure Level.

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

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

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

ANSI/ASA S12.43-1997 (R 2007) American National Standard Methods for Measurement of Sound Emitted by Machinery and Equipment at Workstations and Other Specified Positions.

ANSI/ASA S12.44-1997 (R 2007) American National Standard Methods for Calculation of Sound Emitted by Machinery and Equipment at Workstations and Other Specified Positions from Sound Power Level.

ANSI/ASA S12.50-2002 (R 2007)/ISO 3740:2000 American National Standard Acoustics—Determination of sound power levels of noise sources—Guidelines for the use of basic standards (Nationally Adopted International Standard).

ANSI/ASA S12.51-2002 (R 2007)/ISO 3741:1999 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Precision method for reverberation rooms (Nationally Adopted International Standard). This Standard includes Technical Corrigendum 1-2001. *This standard replaces ANSI S12.31-1990 and ANSI S12.32-1990.*

ANSI S12.53/Part 1-1999 (R 2004)/ISO 3743-1:1994 American National Standard Acoustics—Determination of sound power levels of noise sources—Engineering methods for small, movable sources in reverberant fields—Part 1: Comparison method for hard-walled test rooms (Nationally Adopted International Standard). *This standard, along with ANSI S12.53/Part 2-1999 replaces ANSI S12.33-1990.*

ANSI S12.53/Part 2-1999 (R 2004)/ISO 3743-2:1994 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Engineering methods for small, movable sources in reverberant fields—Part 2: Methods for special reverberation test rooms (Nationally Adopted International Standard). *This standard, along with ANSI S12.53/Part 1-1999 replaces ANSI S12.33-1990.*

ANSI S12.54-1999 (R 2004)/ISO 3744:1994 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Engineering method in an essentially free field over a reflecting plane (Nationally Adopted International Standard). *This standard replaces ANSI S12.34-1988.*

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

ANSI S12.56-1999 (R 2004)/ISO 3746:1995 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Survey method using an enveloping measurement surface over a reflecting plane (Nationally Adopted International Standard). *This standard replaces ANSI S12.36-1990.*

ANSI/ASA S12.57-2002 (R 2007)/ISO 3747:2000 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Comparison method *in situ* (Nationally Adopted International Standard).

ANSI/ASA S12.60-2002 (R 2009) American National Standard Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools.

ANSI/ASA S12.60/Part 2-2009 American National Standard Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools, Part 2: Relocatable Classroom Factors.

ANSI S12.65-2006 American National Standard for Rating Noise with Respect to Speech Interference. *(Revision of ANSI S3.14-1977).*

ANSI/ASA S12.67-2008 American National Standard Pre-Installation Airborne Sound Measurements and Acceptance Criteria of Shipboard Equipment.

ANSI/ASA S12.68-2007 American National Standard Methods of Estimating Effective A-Weighted Sound Pressure Levels When Hearing Protectors are Worn.

ASA Committee on Standards (ASACOS)

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

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

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

ISO/TC 108 Mechanical vibration, shock, and condition monitoring (D.J. Evans, U.S. TAG Chair)

ISO/TC 108/SC 2 Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures (W.C. Foiles and R.F. Taddeo, U.S. TAG Co-Chairs)

ISO/TC 108/SC 3 Use and calibration of vibration and shock measuring instruments (D.J. Evans, U.S. TAG Chair)

ISO/TC 108/SC 4 Human exposure to mechanical vibration and shock (D.D. Reynolds, U.S. TAG Chair)

ISO/TC 108/SC 5 Condition monitoring and diagnostics of machines (D.J. Vendittis, U.S. TAG Chair; R.F. Taddeo, U.S. TAG Vice Chair)

ISO/TC 108/SC 6 Vibration and shock generating systems (C. Peterson, U.S. TAG Chair)

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

Standards News from the United States

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

American National Standards Call for Comment on Proposals Listed

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

AHRI (Air-Conditioning, Heating, and Refrigeration Institute)

New Standards

BSR/AHRI Standard 270-200x, Sound Performance Rating of Outdoor Unitary Equipment (new standard)

Applies to the outdoor sections of factory-made air-conditioning and heat pump equipment as defined in AHRI Standard 210/240 or AHRI Standard 340/360 (cooling capacity ratings of equal to or less than 135,000 Btu/h [40.0 kW]. (August 31, 2009)

BSR/AHRI Standard 300-200x, Sound Rating and Sound Transmission Loss of Packaged Terminal Equipment (new standard)

Applies to the indoor and outdoor sections of factory-made packaged terminal equipment, as defined in AHRI Standard 310/380. (August 31, 2009)

BSR/AHRI Standard 575-200x, Method of Measuring Sound within an Equipment Space (new standard)

Applies to water chilling systems, pumps and similar operating machines and parts thereof, which, for reasons of size or operating characteristics, are more practically evaluated in situ. (August 31, 2009)

ASA (ASC S1) (Acoustical Society of America)

Revisions

BSR/ASA S1.18-200x, Method for Determining the Acoustic Impedance of Ground Surfaces (revision and redesignation of ANSI S1.18-1999 (R2004))

Describes procedures for obtaining the acoustic impedance of ground surfaces from in-situ measurements of the magnitudes and relative phase of the sound pressures at two vertically separated microphones using specified geometries. This standard extends and revises the template method published as ANSI S1.18-1999 to enable the user to obtain impedance spectra that result entirely from measurements and are independent of any model for the acoustic impedance of the ground. (August 24, 2009)

ASA (ASC S3) (Acoustical Society of America)

Revisions

BSR/ASA S3.25-200x, Occluded Ear Simulator (revision and redesignation of ANSI S3.25-1989 (R2003))

Gives acoustical performance criteria for a device that provides acoustic impedance and exhibits sound-pressure distributions approximating the median adult human ear between an earmold and the eardrum. Two specific embodiments whose performance conforms to these criteria are described. As a simulation of part of a median adult human ear, the occluded ear simulator is suitable for use in test systems such as manikins, where the complete ear is to be simulated. (August 24, 2009)

ASA (ASC S2) (Acoustical Society of America)

New National Adoptions

BSR ASA S2.1-200x/ISO 2041-2009, Mechanical vibration, shock and condition monitoring—Vocabulary (identical national adoption and revision of ANSI S2.1-2000/ISO 2041-1990)

Reflects advances in technology and refinements in terms used in the original vocabulary standard. As such, it incorporates more precise definitions of some terms, reflecting changes in accepted meaning, and new terms, which were driven by changes in technology (primarily in the areas of signal processing, condition monitoring and vibration and shock diagnostics and prognostics). (August 31, 2009)

Reaffirmations

BSR/ASA S2.31-1979 (R200x), Methods for the Experimental Determination of Mechanical Mobility—Part 1: Basic Definitions and Transducers (reaffirmation and redesignation of ANSI S2.31-1979 (R2004))

Provides basic definitions with comments and identifies the calibration tests, environmental tests, and physical measurements necessary to determine the suitability of impedance heads, force transducers, and accelerometers fo use in measuring mechanical mobility.

BSR/ASA S2.32-1982 (R200x), Methods for the Experimental Determination of Mechanical Mobility—Part 2: Measurements Using Single-Point Translational Excitation (reaffirmation and redesignation of ANSI S2.32-1982 (R2004))

Includes measurement of mobility, acceleration, or dynamic compliance, either as a driving point measurement or as a transfer measurement. This standard also applies to the determination of the arithmetic reciprocals of those ratios as free effective mass.

ASA (ASC S3) (Acoustical Society of America)

Revisions

BSR/ASA S3.22-200x, Specification of Hearing Aid Characteristics (revision of ANSI/ASA S3.22-2009)

Describes air-conduction hearing-aid measurement methods that are particularly suitable for specification and tolerance purposes. Various test methods are described. Specific configurations are given for measuring the input SPL to a hearing aid. Allowable tolerances in relation to values specified by the manufacturer are given for certain parameters. Correction of error: Text in Subclause 5.1 to be changed: "Atmospheric pressure: 760 (+35,−150) mm of Hg or 101.3 (+5,10−20) kPa" (September 27, 2009)

BSR ASA S3.35-200x, Method of Measurement of Performance Character-

istics of Hearing Aids Under Simulated Real-Ear Working Conditions (revision of ANSI S3.35-2004)
Describes methods to measure the acoustical effects of a simulated median adult wearer on the performance of a hearing aid using: direct simulated real-ear aided measurements (sound pressure developed by a hearing aid in an ear simulator for a given free-field input sound pressure), and insertion measurements (the difference between the sound pressures developed in the ear simulator with and without a hearing aid in place). These test methods are not intended for quality control. (August 31, 2009)

ASA (ASC S12) (Acoustical Society of America)

New Standards
BSR/ASA S12.64-200x, Quantities and Procedures for Description and Measurement of Underwater Sound from Ships—Part 1: General Requirements (new standard)
Describes the measurement systems, procedures, and methodologies used for the beam aspect measurement of underwater sound pressure levels from ships at given operating conditions. Resulting quantities are nominal source level values. Does not require use of specific ocean location, but provides requirements for an ocean test site. Underwater SPL measurements are performed in the far-field and then corrected to a reference distance of 1 m. Applicable to all surface vessels, manned or unmanned. (September 14, 2009)

New National Adoptions
BSR/ASA S12.10-200x/Part 1/ISO 7779:200x MOD, Measurement of Airborne Noise Emitted by Information Technology and Telecommunications Equipment—Part 1: Sound Power Level and Emission Sound Pressure Level (national adoption and revision of ANSI/ASA S12.10-2002/ISO 7779:1999 (R2007) (incl AMD1))
Specifies methods for measurement of airborne noise emitted by information technology and telecommunication equipment. (September 21, 2009)

Revisions
BSR/ASA S12.42-200x, Methods for the Measurement of Insertion Loss of Hearing Protection Devices in Continuous or Impulsive Noise Using Microphone-In-Real-Ear or Acoustic Test Fixture Procedures (revision and redesignation of ANSI/ASA S12.42-1995 (R2004))
Provides two methods for measuring the insertion loss of any hearing protection device (HPD) that encloses the ears, caps the ears, or occludes the ear canals. This standard contains information on instrumentation, calibration, electroacoustic requirements, subject selection and training, procedures for locating ear-mounted microphones and HPDs to measure sound pressure levels at the ear, specifications describing suitable ATFs, and methods for reporting the calculated insertion-loss values. (September 21, 2009)

Reaffirmations
BSR S12.53/Part 2-1999/ISO 3743-2-1994 (R2004), Acoustics—Determination of sound power levels of noise sources using sound pressure—Engineering methods for small, movable sources in reverberant fields—Part 2: Methods for special reverberation test rooms (reaffirmation and redesignation of ANSI S12.53/Part 2-1999/ISO 3743-2-1994 (R2004))
Specifies a relatively simple engineering method for determining the sound power levels of small, movable noise sources.

ASME (American Society of Mechanical Engineers)

Revisions
BSR/ASME BPVC Section V-200x, Nondestructive Examination (revision of ANSI/ASME BPVC 2007 Edition)
Provides requirements and methods for nondestructive examination (NDE). These requirements and methods are intended for use in the Sections of the ASME Boiler and Pressure Vessel Code covering construction of components and items and their integrity in service. These NDE methods are intended to detect surface and internal imperfections in materials, welds, fabricated parts, and components. They include radiographic examination, ultrasonic examination, liquid penetrant examination, magnetic particle examination, eddy current examination, visual examination, leak testing, and acoustic emission examination. (August 31, 2009)

Project Initiation Notification System (PINS)
ANSI Procedures require notification of ANSI by ANSI-accredited standards developers of the initiation and scope of activities expected to result in new or revised American National Standards. This information is a key element in planning and coordinating American National Standards. The following is a list of proposed new American National Standards or revisions to existing American National Standards that have been received from ANSI-accredited standards developers that utilize the periodic maintenance option in connection with their standards. Directly and materially affected interests wishing to receive more information should contact the standards developer directly.

ASA (ASC S3) (Acoustical Society of America)

BSR/ASA S3.22-200x, Specification of Hearing Aid Characteristics (revision of ANSI/ASA S3.22-2009)

Describes air-conduction hearing-aid measurement methods that are particularly suitable for specification and tolerance purposes. Various test methods are described. Specific configurations are given for measuring the input SPL to a hearing aid. Allowable tolerances in relation to values specified by the manufacturer are given for certain parameters. Project Need: To correct a typographical error in subclause 5.1. Stakeholders: Hearing aid manufacturers, hearing aid dispensers.

ASA (ASC S12) (Acoustical Society of America)

BSR/ASA S12.54-20XX/ISO 3744:20XX, Acoustics—Determination of Sound Power Levels and Sound Energy Levels of Noise Sources Using Sound Pressure—Engineering Method for an Essentially Free Field Over a Reflecting Plane (identical national adoption and revision of ANSI S12.54-1999/ISO 3744-1994 (R2004))

Specifies methods for determining the sound power level or sound energy level of a noise source from sound pressure levels measured on a surface enveloping the noise source in an environment that approximates an acoustic free-field near one or more reflecting planes. The sound power level (or, in the case of noise bursts or transient noise emission, the sound energy level) produced by the source, in frequency bands or with frequency-weighting A applied, is calculated using those measurements. Project Need: The current ANS is an identical national adoption. The underlying ISO document is undergoing revision and the new version is expected within the next few months. Upon its publication, it is expected that the new version will be proposed for identical national adoption. Stakeholders: Noise control engineers, manufacturers, researchers.

BSR/ASA S12.56-20XX/ISO 3746:20XX, Acoustics—Determination of Sound Power Levels and Sound Energy Levels of Noise Sources Using Sound Pressure—Survey Method Using an Enveloping Measurement Surface Over a Reflecting Plane (identical national adoption and revision of ANSI S12.56-1999/ISO 3746-1995 (R2004))

Specifies methods for determining the sound power levels of a noise source from sound pressure levels measured on a surface enveloping a noise source (machinery or equipment) in a test environment for which requirements are given. The sound power level (or, in the case of noise bursts or transient noise emission, the sound energy level) produced by the noise source, with frequency-weighting A applied, is calculated using those measurements. Project Need: The current ANS is an identical national adoption. The underlying ISO document is undergoing revision and the new version is expected within the next few months. Upon its publication, it is expected that the new version will be proposed for identical national adoption. Stakeholders: Noise control engineers, manufacturers, researchers.

BSR/ASA S12.53/Part 1-20XX/ISO 3743-1:20XX, Acoustics—Determination of Sound Power Levels and Sound Energy Levels of Noise Sources Using Sound Pressure—Engineering Methods for Small, Movable Sources in Reverberant Fields—Part 1: Comparison Method for Hard-Walled Test Rooms (identical national adoption and revision of ANSI S12.53/Part 1-1999 ISO 3743-1-1994 (R2004))

Specifies methods for determining the sound power level or sound energy level of a noise source by comparing measured sound pressure levels emitted by this source (machinery or equipment) mounted in a hard-walled test room with those from a calibrated reference sound source. The sound power level (or, in the case of noise bursts or transient noise emission, the sound energy level) produced by the noise source is calculated using those measurements. Project Need: The current ANS is an identical national adoption. The underlying ISO document is undergoing revision and the new version is expected within the next few months. Upon its publication, it is expected that the new version will be proposed for identical national adoption. Stakeholders: Noise control engineers, manufacturers, researchers.

ASME (American Society of Mechanical Engineers)

BSR/ASME MFC 5.1M-200x, Measurement of Liquid Flow in Closed Conduits Using Transit-Time Ultrasonic Flowmeters (revision and partition of ANSI/ASME MFC-5M-1985 (R2006))

Applies to ultrasonic flowmeters that base their operation on the measurement of transit time of acoustic signals. This standard concerns the volume flowrate measurement of a single phase liquid with steady flow or flow varying only slowly with time in a completely filled closed conduit. Project Need: To revise and update this standard to reflect the current state of the art. Stakeholders: Manufacturers and users of transit-time ultrasonic flowmeters.

ASSE (ASC A10) (American Society of Safety Engineers)

BSR/ASSE A10.46-200x, Hearing Loss Prevention in Construction and Demolition Workers (revision of ANSI/ASSE A10.46-2007)

Applies to all construction and demolition workers with potential noise exposures (continuous, intermittent and impulse) of 85 dBA and above. Project Need: To make corrections based upon the consensus of the ASC A10—Construction and Demolition Operations. Stakeholders: SH&E Professionals working in the construction and demolition industry.

ASTM (ASTM International)

BSR/ASTM WK25482-200x, New Test Method for Shock Test for Structural Insulation of a Class Divisions Constructed of Steel or Aluminum (new standard)

http://www.astm.org/DATABASE.CART/WORKITEMS/WK25482.htm. Project Need: http:www.astm.org/DATABASE.CART/WORKITEMS/WK25482.htm. Stakeholders: Ships and marine technology industry.

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

BSR/TIA 470.110-D-200x, Telecommunications—Telephone Terminal Equipment—Handset Acoustic Performance Requirements for Analog Telephones (revision and redesignation of ANSI/TIA 470-110-C-2004)

Establishes handset telephone acoustic transmission performance requirements for analog telephones. Corrections and a number of additions will be made, particularly related to telephones designed for the hard-of-hearing. Project Need: To establish handset telephone acoustic transmission performance requirements for analog telephones. Stakeholders: Telecommunications Industry Association.

BSR/TIA 470-220-D-200x, Telecommunications—Telephone Terminal Equipment—Alerter Acoustic Output Performance Requirements for Analog Telephones (revision and redesignation of ANSI/TIA 470.220-C-2004)

Establishes alerter acoustic output performance requirements for analog telephones. Proposed changes include: removing A-weighting; defining a frequency range for the measurements; and including a recommended high output level and frequency spectrum to improve accessibility for people who are hard-of-hearing. Project Need: To establish alerter acoustic output performance requirements for analog telephones. Stakeholders: Telecommunications Industry Association.

TCNA (ASC A108) (Tile Council of North America)

BSR A118.13-200x, Specifications for Bonded Sound Reduction Membranes for Thin-Set Ceramic Tile and Dimension Stone Installation (new standard)

Bonded Sound Reduction membranes for thin-set ceramic tile and dimension stone installations lower the transmission of sound from one room to the room below. Membranes covered by this specification are bonded to a variety of manufacturer-approved substrates covered by ANSI specifications. Project Need: To create a revision that addresses sound reduction membranes. Stakeholders: Ceramic tile installers, contractors, and builders, related material manufacturers, distributors, retailers.

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.

ATIS (Alliance for Telecommunications Industry Solutions)

New Standards

ANSI ATIS 0600010.02-2009, Equipment Handling, Transportation Vibration and Rail Car Shock Requirements for Network Telecommunications Equipment (new standard)

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

Revisions

ANSI C63.4-2009, Methods of Measurement of Radio-Noise Emissions from Low-Voltage Electrical and Electronic Equipment in the Range of 9 kHz to 40 GHz (revision of ANSI C63.4-2003)

InfoComm (InfoComm International)

New Standards

ANSI/INFOCOMM 1M-2009, Audio Coverage Uniformity in Enclosed Listener Areas (new standard)

STANDARDS NEWS FROM ABROAD

Newly Published ISO and IEC Standards

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

ISO Standards

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO 16063-31:2009, Methods for the calibration of vibration and shock transducers—Part 31: Testing of transverse vibration sensitivity

SMALL CRAFT (TC 188)

ISO 14509-3:2009, Small craft—Airborne sound emitted by powered recreational craft—Part 3: Sound assessment using calculation and measurement procedures

IEC Standards

AUDIO, VIDEO AND MULTIMEDIA SYSTEMS AND EQUIPMENT (TC 100)

IEC 60268-17 Ed. 1.0 en Cor.1:1991, Corrigendum 1—Sound system equipment—Part 17: Standard volume indicators

IEC 62106 Ed. 2.0 en:2009, Specification of the Radio Data System (RDS) for VHF/FM sound broadcasting in the frequency range from 87,5 MHz to 108,0 MHz

ELECTRICAL EQUIPMENT IN MEDICAL PRACTICE (TC 62)

IEC 60601-2-5 Ed. 3.0 b:2009, Medical electrical equipment—Part 2-5: Particular requirements for the basic safety and essential performance of ultrasonic physiotherapy equipment

FIBRE OPTICS (TC 86)

IEC 61300-2-1 Ed. 3.0 b:2009, Fibre optic interconnecting devices and passive components—Basic test and measurement procedures—Part 2-1: Tests—Vibration (sinusoidal)

NUCLEAR INSTRUMENTATION (TC 45)

IEC 60988 Ed. 2.0 b:2009, Nuclear power plants—Instrumentation important to safety—Acoustic monitoring systems for detection of loose parts: characteristics, design criteria and operational procedures

PERFORMANCE OF HOUSEHOLD ELECTRICAL APPLIANCES (TC 59)

IEC 60704-2-2 Ed. 2.0 en:2009, Household and similar electrical appliances—Test code for the determination of airborne acoustical noise—Part 2-2: Particular requirements for fan heaters

ISO Draft Standards

BUILDING CONSTRUCTION MACHINERY AND EQUIPMENT (TC 195)

ISO/DIS 18651-1, Building construction machinery and equipment—Internal vibrators for concrete—Part 1: Terminology and commercial specifications—10/31/2009

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO/DIS 29821-1, Condition monitoring and diagnostics of machines—Ultrasound—Part 1: General guidelines—11/21/2009

REVIEWS OF ACOUSTICAL PATENTS

Sean A. Fulop
Dept. of Linguistics, PB92
California State University Fresno
5245 N. Backer Ave., Fresno, California 93740

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

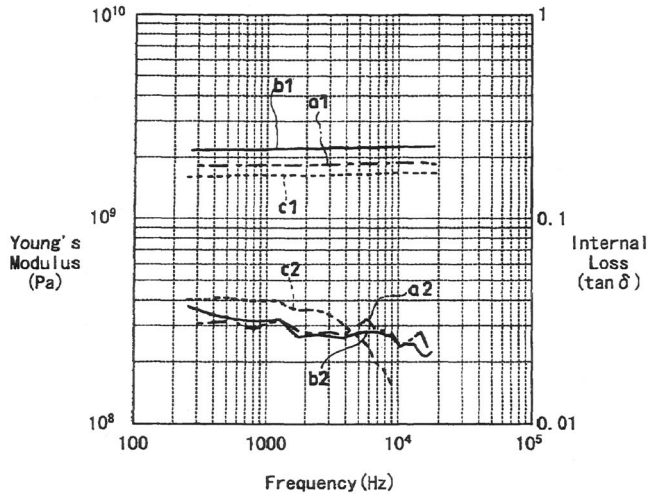
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ANGELO CAMPANELLA, *3201 Ridgewood Drive, Hilliard, Ohio 43026-2453*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden 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*
KEVIN P. SHEPHERD, *Mail Stop 463, NASA Langley Research Center, Hampton, Virginia 23681*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627*

7,510,627

43.38.Ar ACOUSTIC PAPER DIAPHRAGM AND ACOUSTIC TRANSDUCER APPARATUS

Masaru Uryu and Kunihiko Tokura, assignors to Sony Corporation
31 March 2009 (Class 162/138); filed in Japan 19 March 2004

Recycling of electronic components is mandated in many countries. Some materials used in paper cones are not biodegradable, therefore posing end-of-life disposal problems. By using polylactide resin emulsion, which is biodegradable, as a sizing agent in paper cone manufacturing, not only is the product recyclable but, when implemented per the patent, the cone can have increased Young's modulus and better internal loss characteristics as well as better moisture resistance compared to plain pulp and pulp with latex based



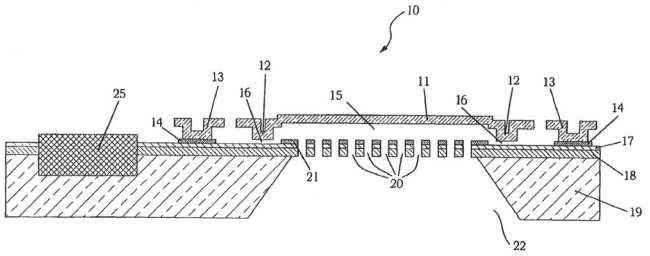
sizing agents. Curves "a" are for plain paper pulp, curves "b" are for pulp internally sized with polylactide resin, and curves "c" are for pulp sized with a latex based agent. Curve "b1" shows higher Young's modulus compared to plain and latex sized pulps, while curve "b2" shows more consistent internal damping compared to the latex sized pulp of curve "c2."—NAS

7,536,769

43.38.Bs METHOD OF FABRICATING AN ACOUSTIC TRANSDUCER

Michael Pedersen, assignor to Corporation for National Research Initiatives
26 May 2009 (Class 29/594); filed 25 May 2006

The authors disclose a method of fabricating a microphone that utilizes relatively simple, traditional micro-electronic mechanical systems fabrication techniques. The microphone is somewhat unusual in that it is directional, thanks to having a set of distributed leaks across the back of the diaphragm, as opposed to a single leak into the back cavity. A novel feature of this design is that it uses additive processes (metal deposition) to build the



diaphragm 11, while the back plate 18 is formed by boron doping the original wafer. It all seems well thought out, but not particularly novel. The technologies involved in this device are probably sufficiently dated that it is only of interest as a research device.—JAH

7,545,075

43.38.Bs CAPACITIVE MICROMACHINED ULTRASONIC TRANSDUCER ARRAY WITH THROUGH-SUBSTRATE ELECTRICAL CONNECTION AND METHOD OF FABRICATING SAME

Yongli Huang *et al.*, assignors to The Board of Trustees of the Leland Stanford Junior University
9 June 2009 (Class 310/309); filed 4 June 2005

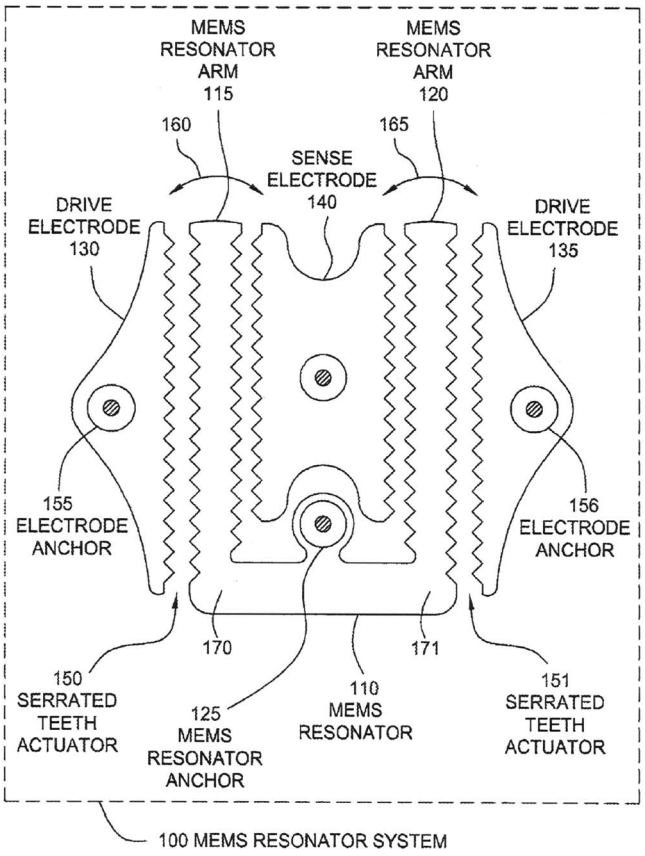
The *capacitive micromachined ultrasonic transducer* was patented about 5 years ago with some basic design calculations that showed useful ultrasonic frequency response and rf impedance characteristics. Building on this, the inventors have made some innovations in the construction details. Whereas the original transducers were planar arrays on a silicon wafer, they now have found ways of back-etching the wafer to make it both flexible enough to wrap around curves while allowing for electrical vias through the wafer so that the support electronics can be stacked beneath it. This latter feature is important for operation of these devices, as they have very little capacitance and so create a rather difficult source impedance to work with.—JAH

7,545,239

43.38.Bs SERRATED MEMS RESONATORS

Paul Merritt Hagelin and David Raymond Pedersen, assignors to SiTime Incorporated
9 June 2009 (Class 333/186); filed 20 December 2006

This patent discloses interesting tradeoffs regarding the use of angular (serrated) electrodes on electrostatic actuators. The authors consider the particular case of a micromachined tuning fork being driven by adjacent electrodes and demonstrate how the linearity of the device response is improved if the faces of the electrodes are serrated as shown in the figure. In this figure, which depicts a tuning fork resonator, the tuning fork bars 160, 165 have triangular serrations on them so as to make the faces meet at an angle to their motion. The patent teaches that this angle is useful as a parameter to control the nonlinearity in the response of the tuning fork to drive voltages on the drive electrodes 130, 135. This is true, but it is not clear why the authors do not go all the way to a comb drive, in which the angles are essentially 90°.—JAH

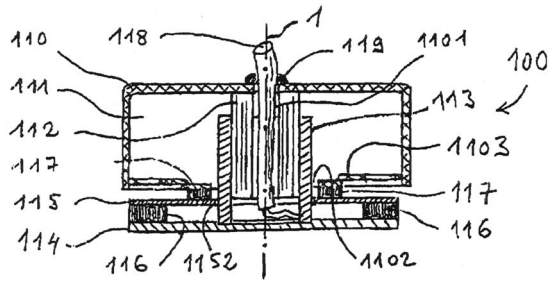


7,545,948

43.38.Dv LOUDSPEAKER

Bernard Fradin, 44350 Guérande, France
9 June 2009 (Class 381/394); filed 1 June 2005

According to the explanatory text of this patent the inventor was working on the design of a physical therapy vibrator when he realized that it would make an ideal audio frequency inertia driver, i.e., a motor assembly that can be attached to table tops, walls, or other resonating surfaces. Induction coil 113 appears to function as a solenoid since there are no magnetic poles indicated and no magnetic return path. The coil is attached to vibration plate 114, which is centered by a second plate 115 and two sets of elastic studs 116 and 117. The design does not seem to have benefited from the very large body of prior art in this field, which includes numerous commercial devices.—GLA

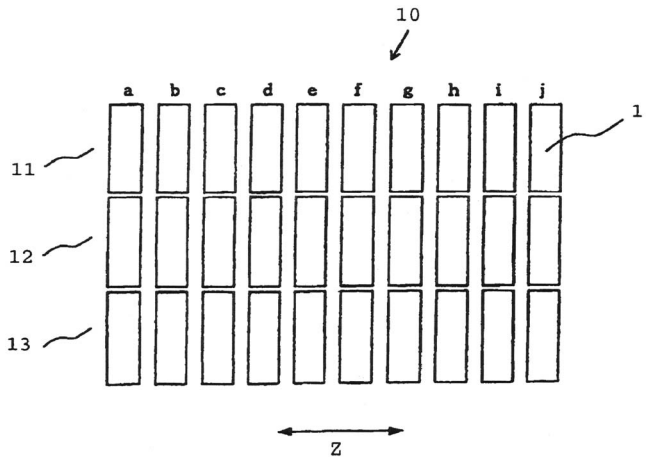


7,443,081

43.38.Hz MULTI-FREQUENCY TRANSMISSION/ RECEPTION APPARATUS

Kazuhiko Kamei *et al.*, assignors to Furuno Electric Company, Limited
28 October 2008 (Class 310/334); filed in Japan 13 April 2001

A design method for ultrasonic transducers with constant directivity over a selected frequency range is claimed, which involves determining the



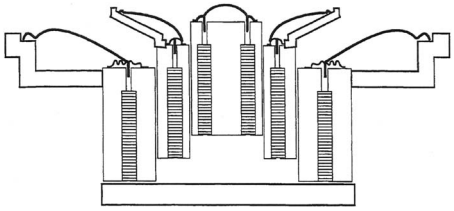
size and shape of a multiplicity of transducer elements assembled into an array. The frequency range includes the higher frequency modes of each element.—AJC

7,515,723

43.38.Hz ELECTRO-ACOUSTIC CONVERTER WITH DEMOUNTABLE DIAPHRAGM AND VOICE COIL ASSEMBLY

Anders Sagren, Uppsala, Sweden
7 April 2009 (Class 381/182); filed 19 July 2006

A co-axial loudspeaker structure is disclosed that allows for adjustment of the drivers relative to one another as well as providing voice-coils,



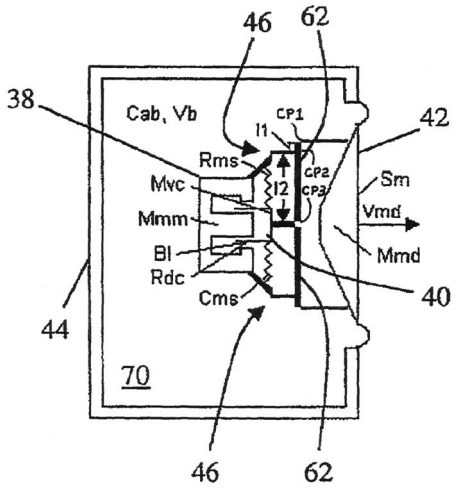
diaphragms, and support elements mounted in self-supporting structures that allow for convenient replacement of same.—NAS

7,508,953

43.38.Ja LOUDSPEAKER AND COMPONENTS FOR USE IN CONSTRUCTION THEREOF

Stefan R. Hlibowicki, assignor to Audio Products International Corporation
24 March 2009 (Class 381/396); filed 30 December 2003

Let us suppose that by converting the vent of a vented enclosure system into a mechanical equivalent, the vent may be disposed of while still retaining a fourth order response. This patent claims this can be done by constructing a fourth order radiating element in which the magnetic motor is not fixed and can move independent of and in addition to the cone, by using



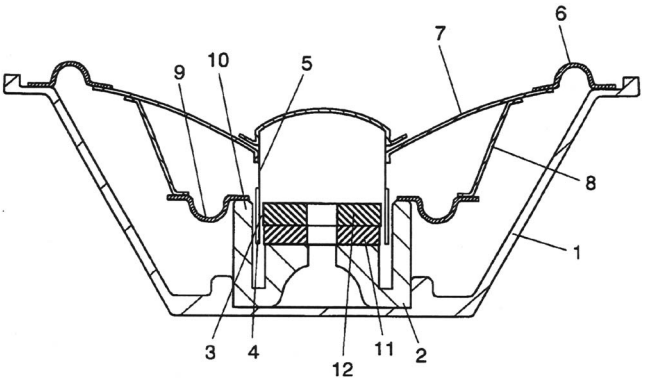
levers (in several variations) between a fixed magnet structure and the cone, or by using both the moving magnet and levers 62 and associated components. The patent contains many analogous circuits, responses, diagrams of embodiments, and a detailed description of these various implementations.—NAS

7,539,323

43.38.Ja SPEAKER

Osamu Funahashi and Seiichi Yoshida, assignors to Panasonic Corporation
26 May 2009 (Class 381/398); filed in Japan 15 March 2005

The loudspeaker shown here is a less complicated version of an carlier design described in U.S. Patent 7,532,736. In both embodiments a second half-roll surround 9 takes the place of a conventional spider (which is called a damper in Japanese patents). While this feature can be expected to allow



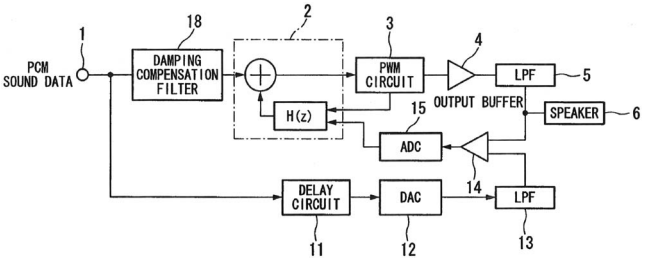
long cone excursions with good mechanical linearity, it has been used in other patented loudspeaker designs.—GLA

7,538,607

43.38.Lc CLASS D AMPLIFIER

Morito Morishima, assignor to Yamaha Corporation
26 May 2009 (Class 330/10); filed in Japan 18 March 2005

The bulk of this patent includes two dozen diagrams and describes 11 embodiments of a single basic concept. In contrast, the patent claims are short and to the point. There is only one independent claim, followed by seven single-sentence variants. Claim 1 describes the circuit shown—a Class D audio power amplifier that is said to provide lower distortion and noise than comparable prior art. The amplifier is driven by a digital audio signal that is transformed to pulse-width modulation (PWM) in block 3. Driver stage 4 and conventional lowpass filter 5 supply an analog signal to



the loudspeaker. Blocks 11, 12, and 13 make up a side chain that creates an "ideal" analog output that is subtracted from the actual output by differential amplifier 14. The resulting error signal is converted back into the digital domain and used to linearize the PWM circuit.—GLA

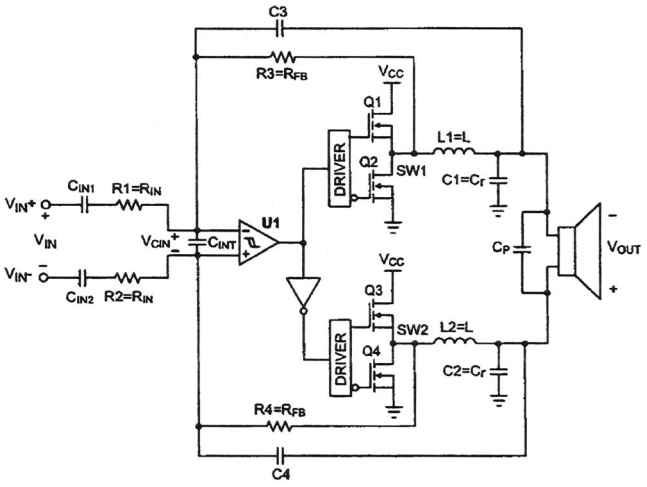
7,538,611

43.38.Lc AUDIO AMPLIFIER WITH HIGH POWER AND HIGH EFFICIENCY

Wei Chen and Peng Xu, assignors to Monolithic Power Systems, Incorporated
26 May 2009 (Class 330/251); filed 7 March 2006

This patent, in contrast to U.S. Patent 7,538,607, consists of a short

descriptive section followed by 20 somewhat wordy claims. The patent describes a bridged version of a "bang-bang" Class D audio power amplifier, which allows the output signal amplitude to reach the full value of the supply voltage. "The Class D amplifier has superior transient response and,



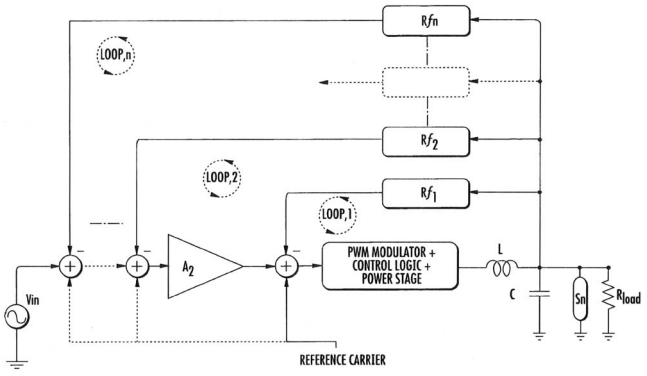
in turn, provides good sound quality and low total harmonic distortion." Two embodiments are described, one using a single comparator and a second using two comparators.—GLA

7,541,869

43.38.Lc FEEDBACK AMPLIFIER

Pietro Mario Adduci and Edoardo Botti, assignors to STMicroelecronics S.r.l.
2 June 2009 (Class 330/251); filed in the European Patent Office 29 March 2006

This is another in a steady stream of patents dealing with distortion reduction in Class D audio amplifiers. In a Class D amplifier, output signal linearity is highly dependent on the actual load impedance, making it difficult to apply significant global feedback without instability. The patent also points out that if a small, iron-core inductor is used in the output demodulation filter (as would be the case in an inexpensive amplifier), its inductance may vary substantially with current, adding another source of nonlinearity.



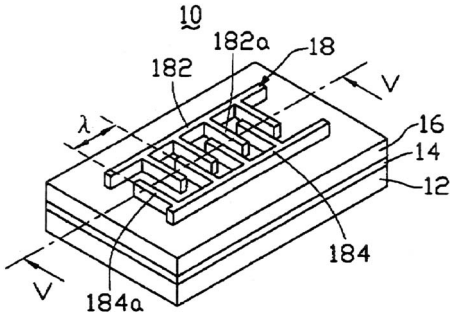
The circuit shown addresses both problems by using multiple feedback loops. Each loop includes a lowpass filter designed to compensate one reactive frequency pole. Performance in the audible band is said to be substantially independent of the output load.—GLA

7,446,452

43.38.Rh SURFACE ACOUSTIC WAVE DEVICE

Ga-Lane Chen, assignor to Hon Hai Precision Industry Company, Limited
4 November 2008 (Class 310/313 R); filed in China 19 October 2005

Surface acoustic wave device 10 operating at frequencies beyond 10 GHz is claimed, comprising interdigital transducer 18-184, a Yo thick zinc oxide piezoelectric layer 16, a Zo thick diamond wave propagation medium 14 (having sound speed more than 10 000 m/s), and substrate 12. In the



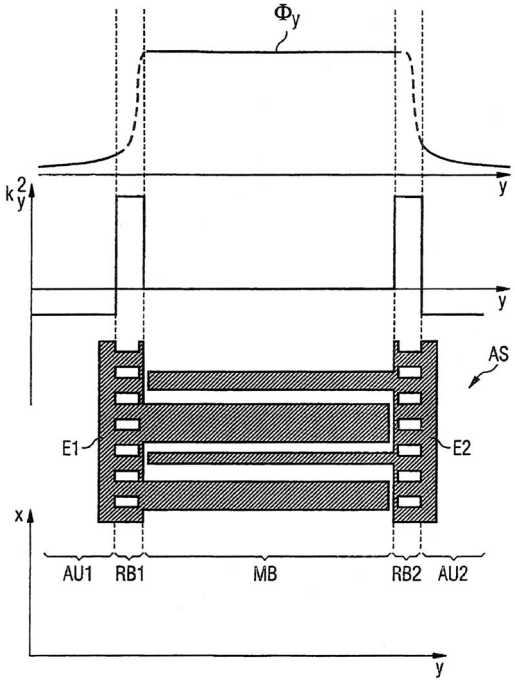
frequency range from 20 to 1000 GHz, the diamond medium thickness ranges from 5 to 20 nm. The zinc oxide piezoelectric layer thickness ranges from 1 to 60 nm. Fabrication means are described.—AJC

7,538,637

43.38.Rh ACOUSTIC WAVE TRANSDUCER WITH TRANSVERSE MODE SUPPRESSION

Markus Mayer *et al.*, assignors to EPCOS AG
26 May 2009 (Class 333/193); filed in Germany 10 July 2003

The authors disclose that a surface acoustic wave device should have "marginal areas" (identified as RB2 in the figure) straddling the usual fingers of the transducer area, where the marginal areas provide a slower phase velocity in which the "edges" of the waves can propagate. The principle



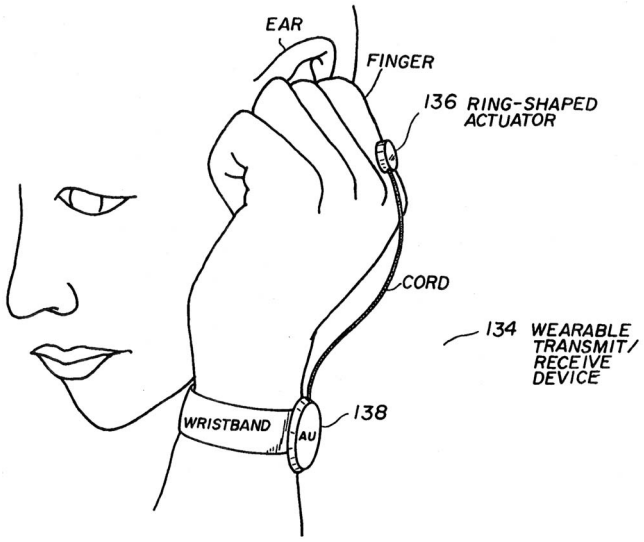
they apply seems to be to control the transverse wave leakage by creating a region with imaginary wave number in the marginal areas. They do this by altering the number of fingers per unit length in that region.—JAH

7,536,020

43.38.Si WEARABLE COMMUNICATION DEVICE

Masaaki Fukumoto and Yoshinobu Tonomura, assignors to Nippon Telegraph and Telephone Corporation
19 May 2009 (Class 381/151); filed in Japan 18 March 1998

This is a continuation of U.S. Patent 6,912,287 filed in 1999. A wearable cellular telephone includes a wristwatch-shaped control module 138, a ring-shaped actuator 136, and a miniature microphone on the inner side of the wristband (not shown). When carrying on a conversation the user blocks his ear with his index finger. The finger also acts as a bone conduction transmitter, energized by vibrations from the actuator. In the earlier patent



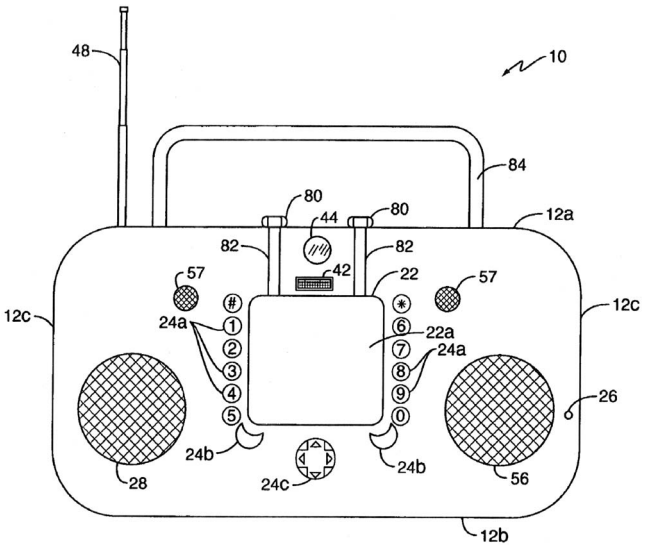
the device was controlled by a tiny keypad on the control module. In this improved version, the user simply taps the control module with his fingernail.—GLA

7,536,200

43.38.Si COMBINATION WIRELESS COMMUNICATIONS DEVICE AND PORTABLE STEREO

William C. Ashman, Jr., assignor to Sony Ericsson Mobile Communications AB
19 May 2009 (Class 455/556.1); filed 22 October 2004

This is one of those patents in which each knob, switch, fastener, and label are listed as a separate claim. The patent describes a portable device that incorporates two loudspeakers 28 and 56, plus microphone 26. When used for voice communication speaker 56 is disabled. Stereo playback utilizes both speakers plus additional speakers 57, whose function is a mystery (tweeters perhaps?). To house all these transducers, one would expect the enclosure to be the size of a lunch pail, yet it is described as "handheld."—GLA

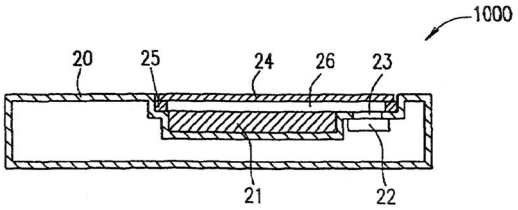


7,536,211

43.38.Si SPEAKER SYSTEM, MOBILE TERMINAL DEVICE, AND ELECTRONIC DEVICE

Shuji Saiki *et al.*, assignors to Panasonic Corporation
19 May 2009 (Class 455/575.1); filed in Japan 28 June 2001

Apart from novelty value, space could be saved by making the video screen of a cellular telephone or laptop computer also function as a sound reproducer. Prior art includes at least a half-dozen patents describing various ways in which this might be done, most of them requiring custom-fabricated displays or complicated driving mechanisms. The patent at hand uses standard components and is easy to implement. A video display panel 21 is viewed through transparent diaphragm 24. The two are separated by a shallow air space. Miniature speaker 22 drives the air space and is thus pneumatically coupled to the diaphragm. The result, not surprisingly, is an acoustic bandpass filter that nonetheless provides reasonably smooth response from about 500 Hz to 10 kHz if various parameters are properly selected.—GLA

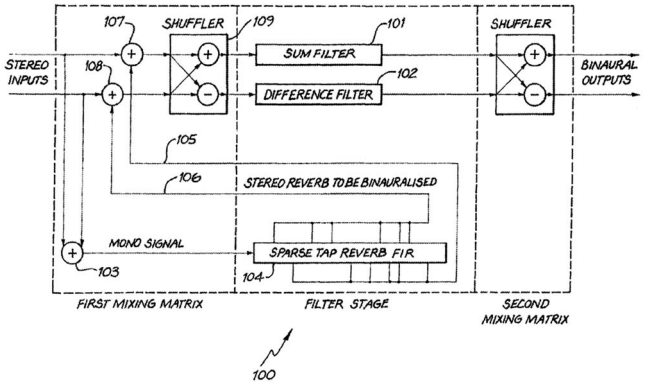


7,536,021

43.38.Vk UTILIZATION OF FILTERING EFFECTS IN STEREO HEADPHONE DEVICES TO ENHANCE SPATIALIZATION OF SOURCE AROUND A LISTENER

Glen Norman Dickins *et al.*, assignors to Dolby Laboratories Licensing Corporation
19 May 2009 (Class 381/310); filed in Australia 16 September 1997

This patent discloses a relatively simple method for processing stereo program material that is to be reproduced through headphones. Simulated reverberation is added to both channels. Sum and difference signals are then



separately filtered to approximate early room reflections. The goal is to create a virtual sound stage surrounding the listener.—GLA

7,539,319

43.38.Vk UTILIZATION OF FILTERING EFFECTS IN STEREO HEADPHONE DEVICES TO ENHANCE SPATIALIZATION OF SOURCE AROUND A LISTENER

Glen Norman Dickins *et al.*, assignors to Dolby Laboratories Licensing Corporation
26 May 2009 (Class 381/310); filed in Australia 16 September 1997

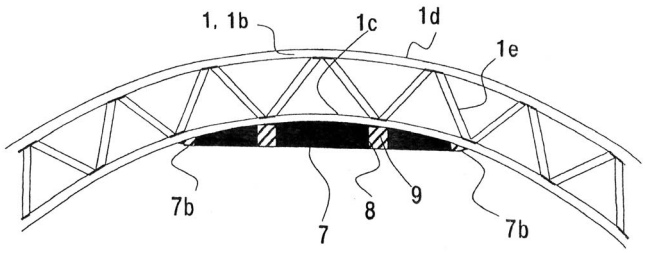
This is a companion patent to U.S. Patent 7,536,021, also reviewed. Here, the same information is described in terms of an apparatus rather than a method.—GLA

7,438,001

43.40.Tm CAR BODY STRUCTURE

Hideyuki Nakamura *et al.*, assignors to Hitachi, Limited
21 October 2008 (Class 105/396); filed in Japan 6 April 2005

An aluminum rail car bending vibration reduction method to improve ride quality is claimed. Roof structure extrusion 1, extending in the direction of the view to the full length of the car, is reinforced by welding formed plate 7 along the center two-thirds of that length. Weldments 7b and 9 through numerous holes 8 are placed along that length. Other embodiments of this vibration reduction method on other car surfaces found along car sides are also claimed.—AJC

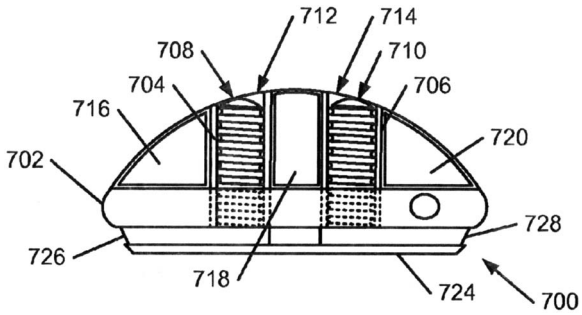


7,510,484

43.40.Tm GOLF CLUB HEAD OR OTHER BALL STRIKING DEVICE WITH MODIFIABLE FEEL CHARACTERISTICS

Gary G. Tavares *et al.*, assignors to Nike, Incorporated
31 March 2009 (Class 473/329); filed 20 February 2008

This is a virtual copy of U.S. Patent 7,354,3559 [reviewed in J. Acoust. Soc. Am. **124**, 2673], in which a golfer is given the ability to modify the "feel" and "give" of a golf club head.—NAS

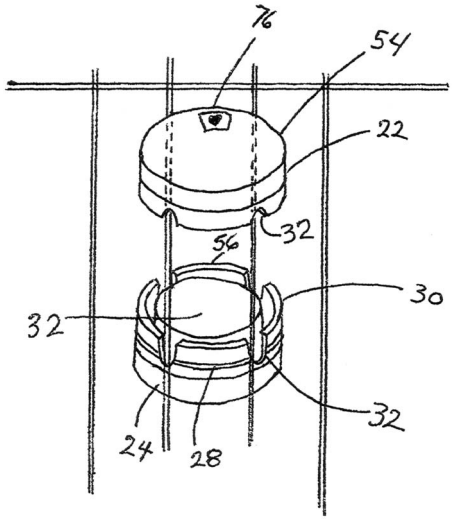


7,517,293

43.40.Tm COMBINATION TENNIS SCORING AND DAMPENING DEVICE

Timothy V. Smith, Cumming, Georgia
14 April 2009 (Class 473/522); filed 17 July 2006

There must be a great need for these combination score and damping units as this is the third patent that this reviewer has seen in the last year. As with U.S. Patents 7,335,118 (reviewed J. Acoust. Soc. Am. **124**, 1900) and 7,427,245 (reviewed J. Acoust. Soc. Am. **125**, 3481), a means is disclosed



for keeping score via rotary dial 54, window 76, and housing member 22, as well as providing for some damping of the strings via the whole assembly and, in this case, especially rubberized material in the core (the left 32 in the figure, not the right 32).—NAS

7,538,273

43.40.Vn CABLE CONNECTION TO DECREASE THE PASSING ON OF VIBRATIONS FROM A FIRST OBJECT TO A SECOND OBJECT

Hans Butler *et al.*, assignors to ASML Netherlands B.V.
26 May 2009 (Class 174/84 R); filed 8 August 2006

In lithographic apparatus and the like, it is desirable to avoid transmission of vibrations to the apparatus via cables that connect the apparatus to external systems. Although loops in the cables provide some of the desired vibration isolation, this may not suffice. This patent describes a means for enhancing this isolation by use of an active control system that minimizes the motion at a suitable point along the cable. A cable support that includes a sensor is attached at that point, and that point's motion is controlled via an actuation arrangement.—EEU

7,533,572

43.40.Yq HIGH BANDWIDTH FIBER OPTIC VIBRATION SENSOR

Michael Twerdochlib, assignor to Siemens Energy, Incorporated
19 May 2009 (Class 73/657); filed 15 August 2006

This sensor is intended for monitoring vibrations in such machines as hydrogen-cooled electric power generators. The optical components are located on the pressurized side of the machine's enclosure so that signals can be fed through the enclosure via electrical cables, thus avoiding the more difficult sealing of fiber-optic cable feed-throughs. The sensor system uses a laser diode, whose beam is partially reflected onto a detector and partially, via a fiber-optic cable, sent to a surface on an elastically supported mass. The light beam reflected from that mass is sent back to the detector via a fiber-optic cable. The resulting electrical signal produced by the detector is processed and fed to an evaluating circuit, which is the only part located outside of the pressurized enclosure.—EEU

7,535,158

43.40.Yq STRESS SENSITIVE ELEMENT

Jun Watanabe, assignor to Epson Toyocom Corporation
19 May 2009 (Class 310/367); filed in Japan 15 February 2007

This patent pertains to an acceleration sensor that consists of a quartz cantilever which carries an end mass and has a doulbe-ended "tuning-fork" type of quartz resonator affixed to one of its surfaces. The cantilever beam bends when it is subjected to acceleration, changing the stress applied to the resonator, and thus its frequency. This frequency change provides a measure of the acceleration. The particular configurations described in the patent are claimed to lend themselves to economical manufacturing.—EEU

7,535,579

43.40.Yq SYSTEM AND METHOD FOR OPTICAL VIBRATION SENSING

Frederick M. Discenzo, assignor to Rockwell Automation Technologies, Incorporated
19 May 2009 (Class 356/498); filed 1 September 2006

Monitoring the vibrations of machines according to this patent is accomplished by measuring the modulation of a light beam which may be reflected from a part of a machine or partially obscured by vibrating machine elements. The light sources may be external or internal to the machine.—EEU

7,536,265

43.40.Yq SIGNAL ANALYSIS METHOD FOR VIBRATORY INTERFEROMETRY

Liang-Chia Chen *et al.*, assignors to Industrial Technology Research Institute
19 May 2009 (Class 702/56); filed in Taiwan 27 March 2007

Optical interferometry has been used widely for static measurement of small (nano-scale) surface profiles. Vibratory measurements have been incorporated in optical interferometric surface profilometers to measure the vibratory behavior of functional elements and thin films in micro-electronic mechanical systems. The present patent describes an approach using a deconvolution operation to determine an object's three-dimensional surface profile from interferometric signals obtained while the test object is operating.—EEU

7,536,910

43.40.Yq VIBRATION ACCELERATION SENSOR

Jun Watanabe, assignor to Epson Toyocom Corporation
26 May 2009 (Class 73/504.12); filed in Japan 20 July 2006

This sensor makes use of two "tuning-fork" piezoelectric vibrating reeds with the same natural frequency oriented on a base so that axial acceleration causes one to be compressed and the other to be in tension. The acceleration is determined from the resulting phase difference between the oscillations of the two reeds, reportedly with high accuracy and little temperature sensitivity.—EEU

7,536,924

43.40.Yq FLEXURE-BASED DYNAMOMETER FOR DETERMINING CUTTING FORCE

Tony Lavaun Schmitz *et al.*, assignors to University of Florida Research Foundation, Incorporated
26 May 2009 (Class 73/862.41); filed 29 September 2005

This device for measuring the cutting forces in high-speed micromachining in essence uses two instrumented sets of flexures in mechanical series. The larger flexure is used to measure the relatively steady cutting forces; the smaller flexure is used to measure the high-frequency oscillations of these forces, which in turn provide an indication of cutting quality and tool wear.—EEU

6,936,000

43.50.Gf PROCESS AND APPARATUS FOR SELECTING OR DESIGNING PRODUCTS HAVING CONTENT-FREE SOUND OUTPUTS

Ray T. Flugger, assignor to Flowmaster, Incorporated
30 August 2005 (Class 600/28); filed 23 May 2003

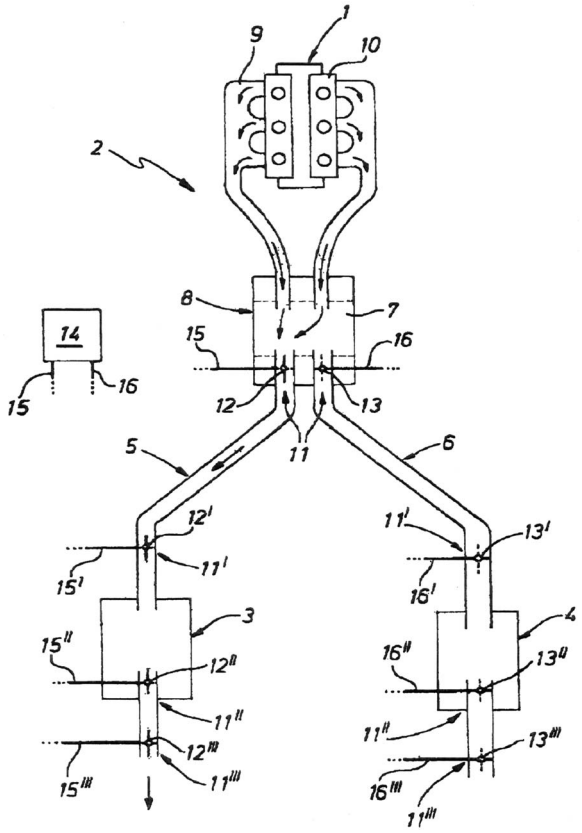
Rather than rely on objective sound measures, comparison of sounds is made via physiological measurements on a test subject. It is claimed that physiological stress measures such as electromyography, electroencephalography, and respiration rate provide a means to determine the most pleasing sounds.—KPS

6,938,729

43.50.Gf EXHAUST GAS SYSTEM

Siegfried Wörner and Ulrich Sigel, assignors to J. Ebersdacher GmbH & Company KG
6 September 2005 (Class 181/254); filed in Germany 10 July 2002

Motor vehicles having two parallel exhaust systems are commonly found with V-engines. A device is proposed that allows switching of the flow paths between the two exhaust systems. This enables, for example, quiet operation at low speeds but reduced backpressure and increased power at higher speeds.—KPS



6,959,956

43.50.Gf ACOUSTICALLY TRANSPARENT VISOR

Mark W. Fero and David J. Prince, assignors to Lear Corporation
1 November 2005 (Class 296/97.5); filed 2 June 2004

A sun visor for use in a motor vehicle is designed to be acoustically transparent to allow sound from a loudspeaker to pass through, and also opaque, so as to inhibit the passage of light. In essence, the design consists of a structural element with many holes covered by fabric. Wow!—KPS

7,540,354

43.50.Gf MICRO-PERFORATED ACOUSTIC LINER

Bruce L. Morin *et al.*, assignors to United Technologies Corporation
2 June 2009 (Class 181/292); filed 26 May 2006

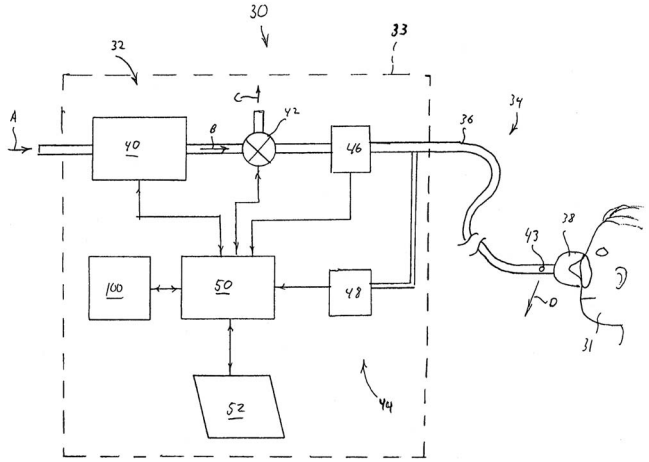
As with many similar sound attenuation liners for a jet engine propulsion system, this one entails a solid backing plate, a cellular honeycomb spacer structure, and a perforated facing. Clever specification of the perforations seems to reduce drag.—CJR

7,448,382

43.50.Ki PRESSURE SUPPORT SYSTEM WITH ACTIVE NOISE CANCELLATION

Paul Alexander *et al.*, assignors to RIC Investments, LLC
11 November 2008 (Class 128/204.18); filed 10 November 2003

A noise suppressor for sleep apnea patient constant positive air pressure system 30 is claimed, combining active noise cancellation 48-100 from a loudspeaker together with an airflow modulation valve 42 to counter noise that may pass through tube 36 to patient 31. The noise is produced by pressure generator 40 which may be a fan, bellows, or piston pump.—AJC



6,933,838

43.50.Lj INFORMING SOUND GENERATION METHOD AND APPARATUS FOR VEHICLE

Kiyoshi Yamaki and Motoaki Miyabe, assignors to Yamaha Corporation
23 August 2005 (Class 340/475); filed in Japan 6 June 2002

A system is designed to inform nearby pedestrians of an automobile driver's intentions. For example, activation of the turn signal would result in sound being generated which is audible to anyone nearby. A rising pitch would indicate a right turn and a falling pitch would indicate a left turn. Similar schemes to indicate acceleration and deceleration would involve changes in the sound's volume, duration, etc. The authors appear unconcerned with the resultant cacophony that would occur on busy streets.—KPS

6,945,677

43.50.Lj MANUAL-OPERATION SOUND AND LIGHT EMITTING DEVICE USED IN VEHICLE

Fon Hsiung Fu, Hsi Hu Chen, and Chang Hua Hsien, Taiwan
20 September 2005 (Class 362/473); filed 20 January 2004

An amazingly complex mechanical device is presented that replaces a traditional bell as found on bicycle handlebars with one that makes sound and light.—KPS

7,537,818

43.55.Ev SOUND ABSORPTIVE MULTILAYER ARTICLES AND METHODS OF PRODUCING SAME

Timothy J. Allison *et al.*, assignors to International Automotive Components Group North America, Incorporated
26 May 2009 (Class 428/95); filed 7 June 2004

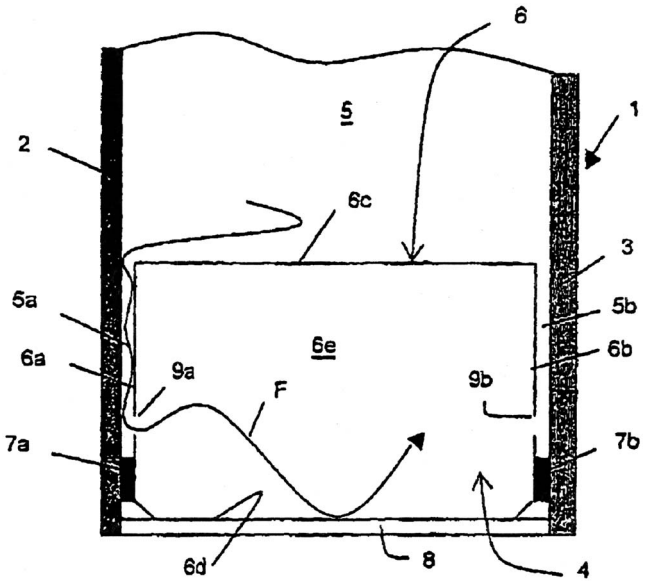
The carpet liner for an automobile cab is enhanced with a porous backing that is applied with a thermo-plastic adhesive.—CJR

7,537,813

43.55.Ti SOUND-INSULATING GLAZING WITH THERMOVISCOUS LOSSES

Beatrice Mottelet and Marc Rehfeld, assignors to Saint-Gobain Glass France
26 May 2009 (Class 428/34); filed in France 15 July 2002

In order to improve the sound isolation performance of an acoustic insulating window, a metal spacer is inserted around the perimeter of the window in such a way as to encourage energy losses from flow resistance into the small cavity created by the spacer.—CJR



7,536,022

43.66.Ts METHOD TO DETERMINE A FEEDBACK THRESHOLD IN A HEARING DEVICE

Andreas Von Buol, assignor to Phonak AG
19 May 2009 (Class 381/318); filed 13 September 2005

A low level input signal is applied to the hearing aid while it is on the wearer to determine the feedback threshold gain of the steady-state closed loop system in multiple frequency bands. Maximum gain in each frequency band is held below the feedback threshold gain for that band.—DAP

7,536,023

43.66.Ts HEARING AID

Marvin A. Leedom *et al.*, assignors to Sarnoff Corporation
19 May 2009 (Class 381/322); filed 17 October 2003

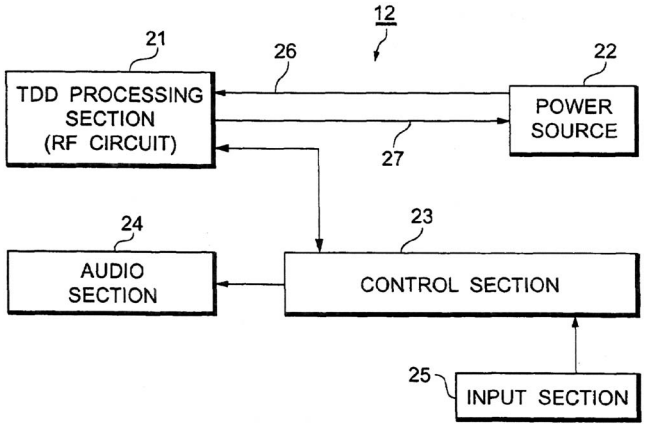
A non-rechargeable power source is permanently integrated into a disposable hearing aid so it cannot be removed. Dependent claims are included for enclosing the hearing aid in a compliant plastic shell and disposing the hearing aid when the power source becomes exhausted.—DAP

7,542,551

43.66.Ts CORDLESS TELEPHONE HANDSET

Kenji Yamazaki and Yoshihisa Takebe, assignors to Uniden Corporation
2 June 2009 (Class 379/52); filed 12 January 2005

The goal is to prevent the current variations in the time division duplex (TDD) processing section of a cordless telephone from causing electromagnetic interference in hearing aids when magnetically coupling the telephone signal to telecoils in hearing aids. When transmitting and receiving in a noise-reduction mode, variations in the magnetic field generated by the TDD processing section of the cordless telephone are suppressed by making the current drawn nearly the same using a power source control circuit and a current smoothing circuit or by canceling the magnetic fields generated by the current.—DAP

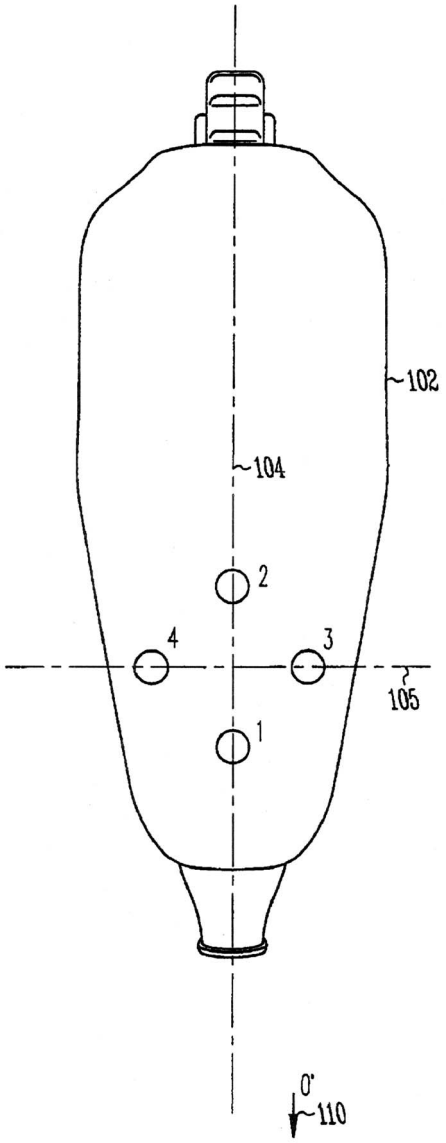


7,542,580

43.66.Ts MICROPHONE PLACEMENT IN HEARING ASSISTANCE DEVICES TO PROVIDE CONTROLLED DIRECTIVITY

Thomas Howard Burns, assignor to Starkey Laboratories, Incorporated
2 June 2009 (Class 381/313); filed 17 July 2006

A first and second pair of directional microphones have intersecting port axes that bisect a frontal view of a hearing aid housing. An omnidirectional microphone may optionally be added. Signal processing electronics in the hearing aid adjust magnitude and phase of the two directional microphone signals and the omnidirectional signal, if present, to provide flexibility in the directional polar patterns achieved.—DAP

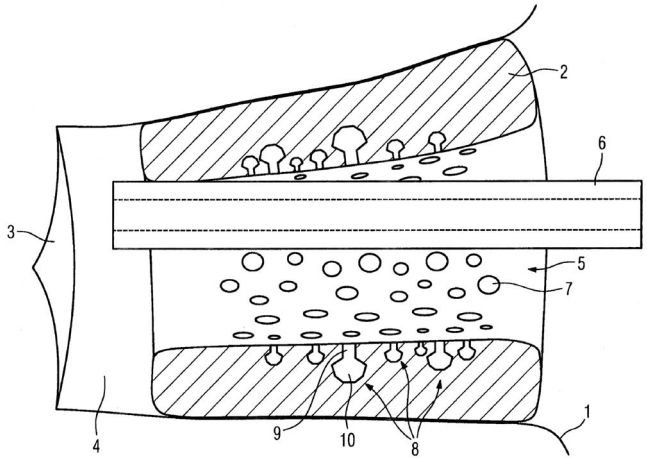


7,542,581

43.66.Ts EAR INSERT FOR A HEARING AID

Joachim Baumann, assignor to Siemens Audiologische Technik GmbH
2 June 2009 (Class 381/328); filed in Germany 5 March 2004

Helmholtz resonators are formed in the hearing aid earmold in order to prevent acoustic feedback problems caused by amplified sounds from a wearer's ear canal getting back to the hearing aid microphone through the earmold vent channel. The Helmholtz resonators are connected physically to the vent channel and are of different sizes to be tuned individually for attenuating different high frequency regions above 1000 Hz.—DAP

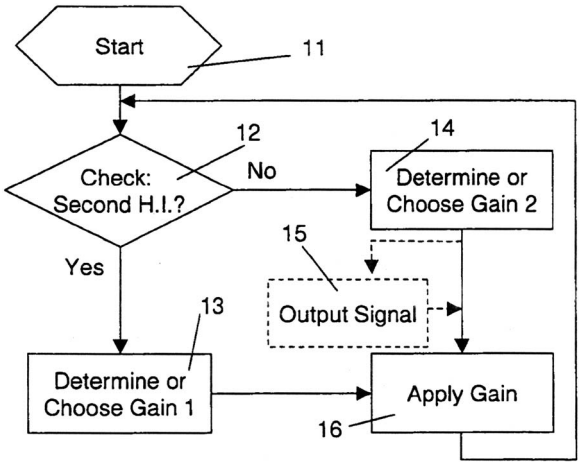


7,545,944

43.66.Ts CONTROLLING A GAIN SETTING IN A HEARING INSTRUMENT

Silvia Allegro Baumann and Stefan Launer, assignors to Phonak AG
9 June 2009 (Class 381/23.1); filed 18 April 2005

A hearing aid determines via wireless or wired communication across the wearer's head whether the other hearing aid in a binaural fitting is present. If so, because of binaural loudness summation, the gain of both hearing aids is set below the gain that would be used for a monaural hearing aid fitting.—DAP



7,542,898

43.72.Gy PITCH CYCLE SEARCH RANGE SETTING APPARATUS AND PITCH CYCLE SEARCH APPARATUS

Kaoru Sato *et al.*, assignors to Panasonic Corporation
2 June 2009 (Class 704/219); filed in Japan 2 August 2001

A speech coding method widely used, for example, in cell phones, is the *code excited linear prediction* method. In order to determine an accurate estimate of the speech pitch value, 20 ms frames of the signal are divided into subframes, typically 5 or 10 ms in length. The speech pitch value is estimated within each subframe and separated into integer and fractional parts. The integer part of the pitch value is used to look up a code book entry which is adapted to the linear prediction values of the speech signal. Exten-

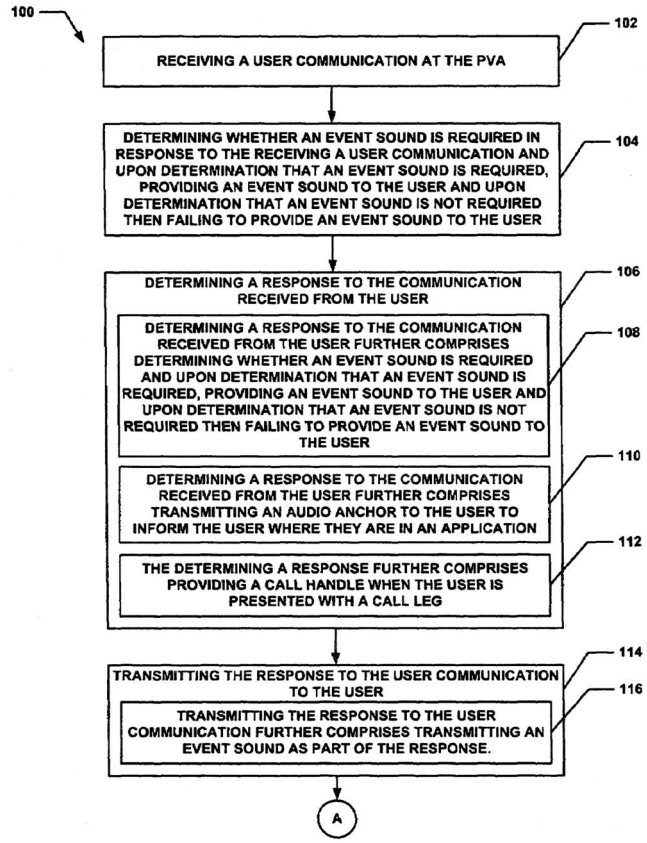
sive detail is provided covering several embodiments of the encoding process.—DLR

7,538,685

43.72.Ne USE OF AUDITORY FEEDBACK AND AUDIO QUEUES IN THE REALIZATION OF A PERSONAL VIRTUAL ASSISTANT

Robert Samuel Cooper *et al.*, assignors to Avaya Incorporated
26 May 2009 (Class 340/692); filed 8 May 2006

The text of this patent includes a detailed and thoughtful discussion of many of the problems and issues interacting with an automated personal assistant. Such issues include things like repetitive or stereotypical behavior of the automated system, lack of understanding of the user's intentions, and inability to follow the progress of the dialog or to keep the user informed of the system's current state. The need to accomplish all of these things through the use of short phrases with proper intonation is stressed. However, nothing is said about the cognitive mechanisms that would be needed to accomplish these lofty goals. A wordy flowchart offers a hint of some of the kinds of details that must be dealt with.—DLR



7,534,210

43.80.Vj METHODS FOR ADAPTIVELY VARYING GAIN DURING ULTRASOUND AGENT QUANTIFICATION

James E. Chomas *et al.*, assignors to Siemens Medical Solution USA, Incorporated
19 May 2009 (Class 600/458); filed 3 February 2004

The gain of an ultrasound system is changed to optimize image intensity based on a triggering event such as the destruction of a contrast agent or the repetition of a contrast agent quantification procedure.—RCW

7,534,239

43.80.Vj AUTOMATIC LIQUID INJECTION SYSTEM AND METHOD

Michel Schneider *et al.*, assignors to Bracco Research S.A.
19 May 2009 (Class 604/500); filed in the European Patent Office 4 December 1997

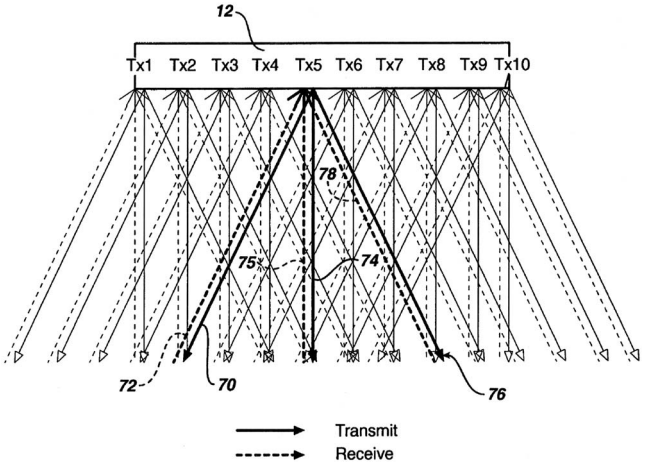
An agent, such as a gas microbubble suspension or lipsone vesicles loaded with iodinated compounds, is controllably delivered with its homogeneity preserved throughout the delivery.—RCW

7,537,567

43.80.Vj ULTRASONIC SPATIAL COMPOUNDING WITH MULTIPLE SIMULTANEOUS BEAM TRANSMISSION

James Jago and Brent Robinson, assignors to Koninklijke Philips Electronics, N.V.
26 May 2009 (Class 600/447); filed 6 August 2004

Ultrasound beams are transmitted in different directions during a common transmit-receive interval. Echoes received from the different transmit beam directions are beamformed in parallel to produce differently steered beams of coherent echo signals. The echoes at the same spatial position from the different beams are combined to produce a spatially compounded image.—RCW



7,540,842

43.80.Vj DIAGNOSTIC ULTRASOUND IMAGING METHOD AND SYSTEM WITH IMPROVED FRAME RATE

David J. Napolitano *et al.*, assignors to Siemens Medical Solutions USA, Incorporated
2 June 2009 (Class 600/443); filed 28 October 2003

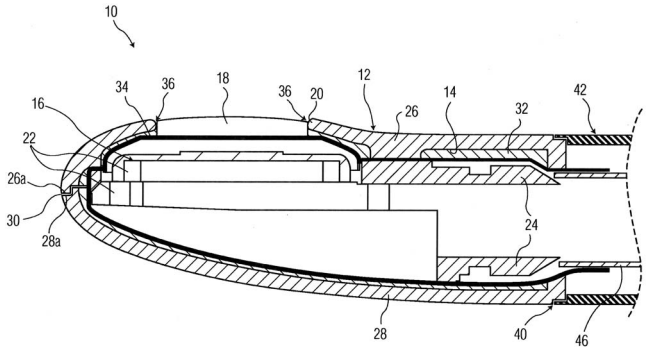
Receive beams that alternate between at least two types of spatially distinct transmit beams are formed. The receive beams associated with each type of transmit beam are then combined. In this way, two-pulse techniques such as phase diversion, synthetic aperture, and synthetic focusing can be implemented by reducing the frame rate penalty normally associated with these techniques.—RCW

7,542,544

43.80.Vj ULTRASOUND GATING OF CARDIAC CT SCANS

Jonathan M. Rubin *et al.*, assignors to The Regents of the University of Michigan
2 June 2009 (Class 378/62); filed 5 January 2005

Cross correlation of ultrasound echo signals from an object such as a moving coronary artery is employed to determine the object location. The method is employed in a prescan procedure to determine an optimal gating window for the acquisition of computer tomographic (CT) image data. The method is also used during the CT scan to determine the gating window in real time.—RCW



7,544,164

43.80.Vj ULTRASOUND PROBES WITH IMPROVED ELECTRICAL ISOLATION

Heather Knowles and Jacquelyn Byron, assignors to Koninklijke Philips Electronics N.V.
9 June 2009 (Class 600/459); filed 4 April 2005

An ultrasound probe containing electrical parts has an acoustic matching layer that electrically isolates the electrical parts from the housing. Seams from between parts of the housing and an acoustic window also serve as electrical barriers. The use of an acoustic matching layer to provide electrical isolation avoids acoustic interference from the electrical barrier.—RCW

7,544,165

43.80.Vj DYNAMICALLY CONFIGURABLE ULTRASOUND TRANSDUCER WITH INTEGRAL BIAS REGULATION AND COMMAND AND CONTROL CIRCUITRY

Donald S. Mamayek *et al.*, assignor to Boston Scientific SciMed, Incorporated
9 June 2009 (Class 600/459); filed 21 October 2004

An array of capacitive transducer elements is connected to a row decoder and a column decoder. The row decoder is coupled to a bias voltage and the column decoder is coupled to a driving signal. A clock is used to synchronize signals between the row decoder and the column decoder.—RCW